

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2003, 20:22:25 ; Search time 104 Seconds  
(without alignments)  
1347.332 Million cell updates/sec

Title: US-09-759-207-2  
Perfect score: 2842  
Sequence: 1 MLRSKALPPLMLLLGP.....LPASYSFVIRNAKVAACI 543

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2838	99.9	543	4	O9Y251
2	2817	99.1	545	4	O9Y251
3	2282	80.3	545	6	O9MYT0
4	2150	75.7	535	11	O8K3K3
5	2123	74.7	536	11	O9QZFB
6	1645.5	57.9	523	13	O9QYK5
7	1154.5	40.6	592	4	O9HB37
8	1146.5	40.3	592	4	O8WMQ2
9	1015.5	35.7	548	4	O8WMQ1
10	936.5	33.0	534	4	O9HB38
11	897.5	31.6	480	4	O9HB39
12	696	24.5	515	5	O8T108
13	416	14.6	521	10	O9SDA1
14	392.5	13.8	543	10	O9EF10
15	381	13.4	527	10	O9LRC8
16					O9LRC8
17					O9LRC8

17	363	12.8	536	10	O9EF21	O9EF21 arabidopsis
18	352.5	12.4	536	10	O9EF21	O9EF21 arabidopsis
19	352.5	12.4	539	10	O8L608	O8L608 arabidopsis
20	169.5	6.0	190	10	O82604	O82604 arabidopsis
21	160	5.6	935	5	O9VE79	O9VE79 drosophila
22	130.5	4.6	493	17	O9HK01	O9HK01 thermoplasma
23	122.5	4.3	408	3	O9HE22	O9HE22 phanerochaete
24	122.5	4.3	408	3	O9HE22	O9HE22 phanerochaete
25	116.5	4.1	617	12	O40996	O40996 meales vir
26	114	4.0	398	16	O8F410	O8F410 leprospira
27	113.5	4.0	617	12	O83295	O83295 meales vir
28	113.5	4.0	1829	2	O9KH44	O9KH44 pantoea agg
29	113	3.9	390	17	O8T108	O8T108 methanococcus
30	112.5	4.0	617	12	O83647	O83647 meales vir
31	111.5	3.9	493	17	O979W0	O979W0 thermoplasma
32	111.5	3.9	575	10	O43855	O43855 vicia faba
33	111.5	3.9	617	12	O40991	O40991 meales vir
34	111.5	3.9	617	12	O9B7H5	O9B7H5 meales vir
35	111	3.9	670	10	O9M090	O9M090 arabidopsis
36	111	3.9	2319	3	O96U00	O96U00 neurospora
37	110.5	3.9	475	5	O8S017	O8S017 encephalito
38	109.5	3.9	617	12	O911P6	O911P6 meales vir
39	109.5	3.9	617	12	O910N9	O910N9 meales vir
40	109.5	3.9	879	16	O8XCP4	O8XCP4 escherichia
41	109	3.8	617	12	O9B7H6	O9B7H6 meales vir
42	108.5	3.8	411	16	P72895	P72895 synchocyst
43	107.5	3.8	617	12	O11381	O11381 meales vir
44	107.5	3.8	617	12	O9A5U0	O9A5U0 cauliobacter
45	107.5	3.8	617	12	O8Q593	O8Q593 meales vir

## ALIGNMENTS

RESULT 1

O9Y251 PRELIMINARY: PRT: 543 AA.

AC O9Y251

DR 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DE HEPRANASE.

GN HPA.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RX MEDLINE=99321249; PubMed=10395326;

RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J., Parish C.R.;

RT "Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis,"

RT Nat. Med. 5:803-809(1999).

RN [2]

RP SEQUENCE FROM N.A.

RA Vlodavsky I., Friedman Y., Elkin M., Aingorn H., Atzmon R., Ishai-Michaeli R., Bitan M., Pappo O., Perez T., Michal I., Spector L., Becker I.;

RT "Mammalian heparanase: a novel gene involved in tumor progression and metastasis,"

RT Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA MEDLINE=99377052; PubMed=10446189;

RX Toyoshima M., Nakajima M.;

RT "Human heparanase. Purification, characterization, cloning, and expression,"

RT J. Biol. Chem. 274:24153-24160(1999).

RN [4]

RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;  
 RA MEDLINE=9935379; PubMed=104051443;  
 RX Seddon A.P., Holmes J.D., Ludwig D., Patel S., Navarro E.C.,  
 RT "Cloning and Functional Expression of a Human Heparanase Gene."  
 RT Biochem. Biophys. Res. Commun. 261:183-187(1999).  
 RL EMBL: AF144325; AAD45379.1;  
 DR EMBL: AF144325; AAD4142.1;  
 DR EMBL: AF155510; AAD54941.1;  
 DR EMBL: AF152376; AAD45669.1;  
 DR Genbank: HGNC:5164; HPSE.  
 DR InterPro: IPR005199; Glyco\_hydro\_79n.  
 DR Pfam: PF03662; Glyco\_hydro\_79n.  
 SQ SEQUENCE 543 AA; 61176 MW; AD262EC67334AB2 CRC64;

Query Match 99.9%; Score 2838; DB 4; Length 543;  
 Best Local Similarity 99.8%; Pred. No. 3,4e-218;  
 Matches 542; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLRSKPLAPPMLLGLGPGALPRPAQADVDLDFPTQEPHLVSPFLSVT	60
DB	1	MLRSKPLAPPMLLGLGPGALPRPAQADVDLDFPTQEPHLVSPFLSVT	60
QY	61	IDANLATDPRFLILGSPKLTARGLSPAYLRFSGTKTDFLIDPKKSTEEERSYMOS	120
DB	61	IDANLATDPRFLILGSPKLTARGLSPAYLRFSGTKTDFLIDPKKSTEEERSYMOS	120
QY	121	QVQVODICKGSIIPPVDEKRLRLEWYQOQLREHYQKKPKSTYSRSSVDVLYTFANCS	180
DB	121	QVQVODICKGSIIPPVDEKRLRLEWYQOQLREHYQKKPKSTYSRSSVDVLYTFANCS	180
QY	181	GDLIFGLNALATADLQNNSSNAQLLDYCSSKGNISWELGNEBNSFLKADIFINCS	240
DB	181	GDLIFGLNALATADLQNNSSNAQLLDYCSSKGNISWELGNEBNSFLKADIFINCS	240
QY	241	QUGEDYIQLHKLKRSKTFKNAKLYGPDVQOPRRKTKMKLSPFKAGGEVYDSVTMHHYVL	300
DB	241	QUGEDYIQLHKLKRSKTFKNAKLYGPDVQOPRRKTKMKLSPFKAGGEVYDSVTMHHYVL	300
QY	301	NGRTATREDPLNDVDLDFISSVQKFFQVVESTRPKKTKWLGETSSAYGGAAPLLSDTA	360
DB	301	NGRTATREDPLNDVDLDFISSVQKFFQVVESTRPKKTKWLGETSSAYGGAAPLLSDTA	360
QY	361	AGFMMLDKLGLSARMGIEVVMROVFPAGAGNYHLVDENFDPPLPDYMLSLFKLVGKTVK	420
DB	361	AGFMMLDKLGLSARMGIEVVMROVFPAGAGNYHLVDENFDPPLPDYMLSLFKLVGKTVK	420
QY	421	ASVQSGSKRRRLRYLHCTNTDNPYKESGDLTYAIVLHNTKYLRLPYFSPNKQVDKYL	480
DB	421	ASVQSGSKRRRLRYLHCTNTDNPYKESGDLTYAIVLHNTKYLRLPYFSPNKQVDKYL	480
QY	481	RPLGPHGLLSKSVQNLGLTKMVDOTLPLMEKPLRPGSSIGLPAFYSFPIYIRNAKVA	540
DB	481	RPLGPHGLLSKSVQNLGLTKMVDOTLPLMEKPLRPGSSIGLPAFYSFPIYIRNAKVA	540
QY	541	ACI 543	
DB	541	ACI 543	

RESULT 2  
 Q9ULJ39 PRELIMINARY; PRT; 545 AA.

AC Q9ULJ39  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Heparanase.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 RN NCBI\_TaxID=9606;  
 RX TISSUE=Placenta;  
 RC

RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA MEDLINE=20229546; PubMed=10764835;  
 RX Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.,  
 RT "Heparanase expression in invasive trophoblasts and acute vascular  
 RT damage."  
 RL Glycobiology 10:467-475(2000).  
 DR EMBL: AF084467; AAD54516.1;  
 DR InterPro: IPR005199; Glyco\_hydro\_79n.  
 DR Pfam: PF03662; Glyco\_hydro\_79n.  
 SQ SEQUENCE 545 AA; 61418 MW; 67B80ACD73C5A9A1 CRC64;

Query Match 99.1%; Score 2817; DB 4; Length 545;  
 Best Local Similarity 99.4%; Pred. No. 1.6e-216;  
 Matches 542; Conservative 1; Mismatches 0; Indels 2; Gaps 2;

QY	1	MLRSKPLAPPMLLGLGPGALPRPAQADVDLDFPTQEPHLVSPFLSVT	58
DB	1	MLRSKPLAPPMLLGLGPGALPRPAQADVDLDFPTQEPHLVSPFLSVT	60
QY	59	VTIDANLATDPRFLILGSPKLTARGLSPAYLRFSGTKTDFLIDPKKSTEEERSYM	118
DB	61	VTIDANLATDPRFLILGSPKLTARGLSPAYLRFSGTKTDFLIDPKKSTEEERSYM	120
QY	119	QVQVODICKGSIIPPVDEKRLRLEWYQOQLREHYQKKPKSTYSRSSVDVLYTFAN	178
DB	121	QVQVODICKGSIIPPVDEKRLRLEWYQOQLREHYQKKPKSTYSRSSVDVLYTFAN	180
QY	179	CSGLDIFGLNALATADLQNNSSNAQLLDYCSSKGNISWELGNEBNSFLKADIFIN	238
DB	181	CSGLDIFGLNALATADLQNNSSNAQLLDYCSSKGNISWELGNEBNSFLKADIFIN	240
QY	239	GSQUGEDYIQLHKLKRSKTFKNAKLYGPDVQOPRRKTKMKLSPFKAGGEVYDSVTMHHY	298
DB	241	GSQUGEDYIQLHKLKRSKTFKNAKLYGPDVQOPRRKTKMKLSPFKAGGEVYDSVTMHHY	300
QY	299	YNGRTATREDPLNDVDLDFISSVQKFFQVVESTRPKKTKWLGETSSAYGGAAPLLSDT	358
DB	301	YNGRTATREDPLNDVDLDFISSVQKFFQVVESTRPKKTKWLGETSSAYGGAAPLLSDT	360
QY	359	FAAGFMMLDKLGLSARMGIEVVMROVFPAGAGNYHLVDENFDPPLPDYMLSLFKLVGKTV	418
DB	361	FAAGFMMLDKLGLSARMGIEVVMROVFPAGAGNYHLVDENFDPPLPDYMLSLFKLVGKTV	420
QY	419	IMAASVQSGSKRRRLRYLHCTNTDNPYKESGDLTYAIVLHNTKYLRLPYFSPNKQVDKYL	478
DB	421	IMAASVQSGSKRRRLRYLHCTNTDNPYKESGDLTYAIVLHNTKYLRLPYFSPNKQVDKYL	480
QY	479	LRLPGLPHGLLSKSVQNLGLTKMVDOTLPLMEKPLRPGSSIGLPAFYSFPIYIRNAK	538
DB	481	LRLPGLPHGLLSKSVQNLGLTKMVDOTLPLMEKPLRPGSSIGLPAFYSFPIYIRNAK	540
QY	539	VAACT 543	
DB	541	VAACT 545	

RESULT 3  
 Q9MYT0 PRELIMINARY; PRT; 545 AA.

AC Q9MYT0  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-JUN-2002 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Heparanase.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 RN NCBI\_TaxID=9913;  
 RX TISSUE=Placenta;  
 RC

RA Kizaki K., Nakano H., Takahashi T., Imai K., Hashizume K.,  
RT "Expression of Heparanase mRNA in Bovine Placenta During Gestation."  
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL/AF281160; AF281160; 2.  
DR InterPro: IPR005199; Glyco\_hydro\_79n.  
DR Pfam: PF03662; Glyco\_hydro\_79n; 1.  
DR Sequence 545 AA; 61076 MW; AACABDFD855B933 CRC64;

Query Match 80.3%; Score 2282; DB 6; Length 545;  
Best Local Similarity 79.8%; Pred. No. 9.5e-174;  
Matches 435; Conservative 35; Mismatches 73; Indels 2; Gaps 1;

DB 1 MLARSRALPPPLML-LLGPGPLSPALPPPAQADVVDLFTQEPHLVSSFLS 58  
1 MLARSRALPPPLML-LLGPGPLSPALPPPAQADVVDLFTQEPHLVSSFLS 58  
DB 59 VTIDANLATPRLPILLLSGPKRTIARGLSPALRFGTITDPLIDPKKESTFEERSY 118  
61 FTIDANLATPRLPILLLSGPKRTIARGLSPALRFGTITDPLIDPKKESTFEERSY 120  
DB 119 QSQVNDICCKYSIPDVEEKLRLMPYEQOLLREHYOKKFNSTYSRSVDVLTYPAN 178  
121 LQSNODICCKYSIPDVEEKLRLMPYEQOLLREHYOKKFNSTYSRSVDVLTYPAN 180  
DB 179 CGLDLIFGLNALIRTDLOMSSNAQLLDYSSKGINISWELGNEPNSFLKADIIFIN 238  
181 CGLDLIFGLNALIRTDLOMSSNAQLLDYSSKGINISWELGNEPNSFLKADIIFIN 240  
DB 239 GQLGGRDYIOLHLKLRSTFKNAKLYGPDVGOPRRKAKMLKSLKAGGEVIDSVTHHY 298  
241 GQLGGRDYIOLHLKLRSTFKNAKLYGPDVGOPRRKAKMLKSLKAGGEVIDSVTHHY 300  
DB 299 YLNGRTATREDPLNPVLDLFISSVQKVFQVNESTRPGKVMLGERTSSAYGGAAPLLSDT 358  
301 YLNGRTATREDPLNPVLDLFISSVQKVFQVNESTRPGKVMLGERTSSAYGGAAPLLSDT 360  
DB 359 PAAGFMWLDLGLSARNGIEVVRQVFFGAGNHLVDENFDPLDYWLSLFFKLVGTXY 418  
361 PAAGFMWLDLGLSARNGIEVVRQVFFGAGNHLVDENFDPLDYWLSLFFKLVGTXY 420  
DB 419 LMSVQSKRRKRLRVYLHCTNTNTPRKYEGDPLTYAINTLVTKYLRLPYFNSKQVDKY 478  
421 LMSVQSKRRKRLRVYLHCTNTNTPRKYEGDPLTYAINTLVTKYLRLPYFNSKQVDKY 480  
DB 479 LRLPLGPHGLSLKSVQNLGLTKLVNDDQTLPLMEKELRPSSIGLPAFSYFFVIRNAK 538  
481 LRLPLGPHGLSLKSVQNLGLTKLVNDDQTLPLMEKELRPSSIGLPAFSYFFVIRNAK 540  
DB 539 VAACT 543  
541 VAACT 545

RESULT 4  
08K33 PRELIMINARY; PRT; 535 AA.

AC 08K33;  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Heparanase.  
OS Mus musculus (Mouse).  
OC Mammalia; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
ON NCBI\_TaxID=10090;  
OX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB;  
RA Maio H.-O., Navarro B., Patel S., Sargent D., Koo H., Wan H.,  
RT "Cloning, expression, and purification of mouse heparanase."  
RL Protein Expr. Purif. 0:0-0(2002).

RN [12]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CS7BL/6J; TISSUE=Thymus;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs."  
RL Nature 420:563-573(2002)  
DR EMBL/AY077467; AAL76083.1;  
DR EMBL/AK040471; BAC30600.1;  
DR MGD: MGI:1343124; Hsae.  
DR InterPro: IPR005199; Glyco\_hydro\_79n.  
DR Pfam: PF03662; Glyco\_hydro\_79n; 1.  
DR Sequence 535 AA; 60065 MW; 6873A8302FB8A0DF CRC64;

Query Match 75.7%; Score 2150; DB 11; Length 535;  
Best Local Similarity 76.6%; Pred. No. 3.3e-163;  
Matches 407; Conservative 51; Mismatches 73; Indels 0; Gaps 0;

DB 13 LMLLIGPLSPALPPPAQADVVDLFTQEPHLVSSFLSVITIDANLATDPRFL 72  
5 LMLLIGPLSPALPPPAQADVVDLFTQEPHLVSSFLSVITIDANLATDPRFL 64  
DB 73 ILLGSPKRTIARGLSPALRFGTITDPLIDPKKESTFEERSYQSQVNDICCKYSI 132  
65 ILLGSPKRTIARGLSPALRFGTITDPLIDPKKESTFEERSYQSQVNDICCKYSI 124  
DB 133 PPVEEKLRLMPYEQOLLREHYOKKFNSTYSRSVDVLTYPANCSGLDLIFGLNAL 192  
125 SAVALRKQVNEPQELLRLRQYOKKFNSTYSRSVDVLTYPANCSGLDLIFGLNAL 184  
DB 193 RTADLOMSSNAQLLDYSSKGINISWELGNEPNSFLKADIIFINGSLGSDYIOLHL 252  
185 RTADLOMSSNAQLLDYSSKGINISWELGNEPNSFLKADIIFINGSLGSDYIOLHL 244  
DB 253 LRSTFKNAKLYGPDVGOPRRKAKMLKSLKAGGEVIDSVTHHYLNGRTATREDPLN 312  
245 LRSTFKNAKLYGPDVGOPRRKAKMLKSLKAGGEVIDSVTHHYLNGRTATREDPLN 304  
DB 313 PVLDLFISSVQKVFQVNESTRPGKVMLGERTSSAYGGAAPLLSDTFAAGFMWLDLGLS 372  
305 PVLDLFISSVQKVFQVNESTRPGKVMLGERTSSAYGGAAPLLSDTFAAGFMWLDLGLS 364  
DB 373 ARNGIEVVRQVFFGAGNHLVDENFDPLDYWLSLFFKLVGTXYLMSVQSKRRKRL 432  
365 ARNGIEVVRQVFFGAGNHLVDENFDPLDYWLSLFFKLVGTXYLMSVQSKRRKRL 424  
DB 433 VYLHCTNTNTPRKYEGDPLTYAINTLVTKYLRLPYFNSKQVDKYLRPLGPHGLSLK 492  
425 VYLHCTNTNTPRKYEGDPLTYAINTLVTKYLRLPYFNSKQVDKYLRPLGPHGLSLK 484  
DB 493 VOLNGLTKLVNDDQTLPLMEKELRPSSIGLPAFSYFFVIRNAKVAACI 543  
485 VOLNGLTKLVNDDQTLPLMEKELRPSSIGLPAFSYFFVIRNAKVAACI 535

RESULT 5  
09QZF8 PRELIMINARY; PRT; 536 AA.

AC 09QZF8;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Heparanase.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
ON NCBI\_TaxID=10116;  
OX [1]  
RP SEQUENCE FROM N.A.  
RA Podjany K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;

RT "Heparanase from parathyroid cell line";

RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.

DR EMBL, AF149667, AAF04563.1, -

DR InterPro, IPR005199, Glyco\_hydro\_79N.

DR Pfam, PF03662, Glyco\_hydro\_79n, 1.

SO SEQUENCE 536 AA; 60569 MW; 6208B1FD9E28421 CRC64;

Query Match 74.7%; Score 2123; DB 11; Length 536;

Best Local Similarity 75.7%; Pred. No. 4.7e-161;

Matches 405; Conservative 51; Mismatches 79; Indels 0; Gaps 0;

```

QY 9 LPPPLMLLLGLPLSPGALPRPAQADVDVLDPFQOEPLHLVSPSLVTIDANLATD 68
  2 LRPILLMLMGRALALVGTSTAGTAFTKDVLDLERYTRQLFQSSPSFSLTTIDASLATD 61
  69 PRFLILGSPRLRLARGLSPAYLRFGGTKTDPLIFDPKKESTFEERSYQSQVNDICK 128
  62 PRFLTFGLSPRLRLARGLSPAYLRFGGTKTDPLIFDPKKESTFEERSYQSQVNDICK 121
  129 YGSLPPVVEKRLAEWPYQGLLREHYQKFKSTYSRSSVDVLYTFANCSGLDILFGL 188
  122 SERVSADVLRQLQEWEPQGLLRLREYQKFKSTYSRSSVDVLYTFANCSGLDILFGL 181
  189 NALRTADLQWSSNAQLLDYCSSKGYNISWEIGNEPNSFLKKADIFINGSLQGEYIQ 248
  182 NALRTADLQWSSNAQLLDYCSSKGYNISWEIGNEPNSFLKKADIFINGSLQGEYIQ 241
  249 LHLKLRSTFPAKALYGPVGOVPRRTAKMLKSFLLKAGEVIDSVTHHYLYNGRTATRE 308
  242 LHLKLRSTFPAKALYGPVGOVPRRTAKMLKSFLLKAGEVIDSVTHHYLYNGRTATRE 301
  309 DELNPVLDIFISSVQKVFQVVESTRPGKKWLGETSAYGGAPLISDTFAAGFMMLDK 368
  302 DFLSSDVLDIFISSVQKVFQVVESTRPGKKWLGETSAYGGAPLISDTFAAGFMMLDK 361
  369 LGLSARWGLEIYVNRQVFFGAGNYHLDENFDPLPDYMLSLFKLVGTGYTLMAVSQSKR 428
  362 LGLSARWGLEIYVNRQVFFGAGNYHLDENFDPLPDYMLSLFKLVGTGYTLMAVSQSKR 421
  429 RKLRYVLTCTNDPRYKGGDLTVAINLHNTKYRLPYPSNKOVDKYLRLPLGPHGL 488
  422 SKLRVYLCTNDPRYKGGDLTVAINLHNTKYRLPYPSNKOVDKYLRLPLGPHGL 481
  489 LSKSVQNLGLTLKAVDDOTLPLMEKPLRPSSSLGLPARFSFVIRNAKAAICI 543
  482 LSKSVQNLGLTLKAVDDOTLPLMEKPLRPSSSLGLPARFSFVIRNAKAAICI 536

```

RESULT 6

ID Q90YK5 PRELIMINARY; PRT; 523 AA.

AC Q90YK5;

DT 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Heparanase.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RX PubMed=11387326;

RA Goldsmith O., Zcharia E., Alingorn H., Guatta-Rangini Z., Atzmon R.,

RA Michael I., Pecker I., Mitrani E., Vlodavsky I.;

RT "Expression Pattern and Secretion of Human and Chicken Heparanase Are

RT Determined by Their Signal Peptide Sequence.";

RL J. Biol. Chem. 276:29178-29187(2001).

DR EMBL, AY037007, AAK82648.1, -

DR InterPro, IPR005199, Glyco\_hydro\_79N.

DR Pfam, PF03662, Glyco\_hydro\_79n, 1.

SO SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;

Query Match 57.9%; Score 1645.5; DB 13; Length 523;

Best Local Similarity 60.2%; Pred. No. 6.7e-123;

Matches 320; Conservative 87; Mismatches 114; Indels 11; Gaps 3;

```

QY 13 LMLLLGLPLSPGALPRPAQADVDVLDPFQOEPLHLVSPSLVTIDANLATDPRFL 72
  2 LVLLLVLLAVLP-----RTTAEQLGIREPIGAVSPALSLTIDASLARPRV 52
  73 ILGSPKRLTLAGLSPAYLRFGGTKTDPLIFDPKKESTFEERSYQSQVNDICKYGS 132
  53 ALLRHKLHTLAGLSPAYLRFGGTKTDPLIFDPKKESTFEERSYQSQVNDICKYGS 111
  133 PPVEEKRLAEWPYQGLLREHYQKFKSTYSRSSVDVLYTFANCSGLDILFGLNALL 192
  112 FAVPPLLLTQWLOELKLLAEHSWKHKNTYTRSTLILHFASSGFRVLFGLNALL 171
  193 RTADLQWSSNAQLLDYCSSKGYNISWEIGNEPNSFLKKADIFINGSLQGEYIQHLK 252
  172 RRGGLQWSSNAQLLDYCSSKGYNISWEIGNEPNSFLKKADIFINGSLQGEYIQHLK 231
  253 L-RKSTFPAKALYGPVGOVPRRTAKMLKSFLLKAGEVIDSVTHHYLYNGRTATRE 311
  232 LSGHPLRYRAELVGLDVQVPRKTHQLRSFKMSGKALDSVTWHHYLYNGRTATRE 291
  312 NPDVLDIFISSVQKVFQVVESTRPGKKWLGETSAYGGAPLISDTFAAGFMMLDK 371
  292 SPVLDLSPFAIADHDVIGVATVPGKKWLGETSAYGGAPLISDTFAAGFMMLDK 351
  372 SARWGLEIYVNRQVFFGAGNYHLDENFDPLPDYMLSLFKLVGTGYTLMAVSQSKR 431
  352 AARGLDVMNRQVFFGAGNYHLDENFDPLPDYMLSLFKLVGTGYTLMAVSQSKR 411
  432 RYVLTCTNDPRYKGGDLTVAINLHNTKYRLPYPSNKOVDKYLRLPLGPHGLSK 491
  412 RYVLTCTNDPRYKGGDLTVAINLHNTKYRLPYPSNKOVDKYLRLPLGPHGLSK 471
  492 SYNLNGLTLKAVDDOTLPLMEKPLRPSSSLGLPARFSFVIRNAKAAICI 543
  472 EYVNLNGLTLKAVDDOTLPLMEKPLRPSSSLGLPARFSFVIRNAKAAICI 523

```

RESULT 7

ID Q9HB37 PRELIMINARY; PRT; 592 AA.

AC Q9HB37;

DT 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Heparanase-like protein HP42C.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20483645; PubMed=11027606;

RA McKenzie B., Tyson K., Stamps A., Smith P., Turner P., Barry R.,

RA Hickey M., Patel S., Barry E., Stuberfield C., Terrett J., Page M.;

RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian

RT Heparanase Family Member.";

RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).

DR EMBL, AF282887, AAG23423.1, -

DR InterPro, IPR005199, Glyco\_hydro\_79N.

DR Pfam, PF03662, Glyco\_hydro\_79n, 1.

SO SEQUENCE 592 AA; 66580 MW; 95C384DA9A74258E CRC64;

Query Match 40.6%; Score 1154.5; DB 4; Length 592;

Best Local Similarity 43.6%; Pred. No. 1.5e-83;

Matches 250; Conservative 82; Mismatches 189; Indels 53; Gaps 9;

```

QY 20 PLGLSPGAL-----PRPA-----QADVDVLDPFQOEPLHLVSPS 55
  1 LGLSPGAL-----PRPA-----QADVDVLDPFQOEPLHLVSPS 55

```

```

Db 18 PPACIAPGALVYALALLHLHLSLSSOAGDRRLPVDBAAGLKEKTLILLDSTKRPVTVNEN 77
Qy 56 FLVSTIDANLADPRLILLGSPKRLTLAGLSPAYLRFGGTDTPLIF---DPKEST 111
Db 78 FLSTLQDPSIIHD-CWLDPLSSKRLVTLARGLSPALRFGKRTDPLQFONLRNPAKSRG 136
Qy 112 FEERSYMSQVUNODI-----CKYGSIPPDVEKRLLEMPYOEOL-LIREHYOK 158
Db 137 GGPDPYLLKNVEDDIIVRSVDVALDKQKCKIAQ-HPDVMLLEQREKAQOMHLVLEKEPSN 195
Qy 159 KFKNSTYSSVDVLYTFPANCGLDLIFGLNALRLTADLQNNSSNAQLLDYCSSKGYNI 218
Db 196 TYSNLIILARSLDKLYNFPADCGSLHLIFALNALRRPNNSNSSLILKYSAKKYNI 255
Qy 219 SWEIENEPNSFLKADIFINGSQJGEDIYQLHKLRLK-STFKNAKLYGPDVGQPRRTAK 277
Db 256 SWEIENEPNNRTYMGRAVNGSQJGKDYIQLKSLQPIRYSRSLYGPNGRBRKNVIA 315
Qy 278 MLKSFILKAGGEVIDSYMHYLLNGRTAREDFLNPVDLIFISSVOKVFOVVESTRPGK 337
Db 316 LLDGPMKAVGSTVDAVTMOHCYIDGRVVKVMDPLKTRLLDLSDOIIRKIQKVNTYTPGK 375
Qy 338 KWLGETSSAYGCGAPLSDTFAAGFMWLDKLSARMGIEVVRQVFFGAGNYHLVDEN 397
Db 376 KIMLEGVTTTSAAGTNNLSDSYAAGFLMNTLGMLANGIDIVIRHSFFDHGNYHLVDON 435
Qy 398 FDPPLDYLWLSLFLKLVGTGKVLMAVSOGSKR-----KLAVYLCTNTDNPRIYEG 448
Db 436 FNPPLDYLWLSLFLKLVGTGKVLMAVSOGSKR-----KLAVYLCTNTDNPRIYEG 495
Qy 449 DLTLYAINLHVNTKYLRLPYFPFSNKQYDKYLLRPLGPHGLSKSVQNLGLTKVWDDOTL 508
Db 496 SITPLTINLHRSRKKIKIAGTLRDKLWQYLLQPYQGEGLSKSVQNLGLTKVWDDOTL 555
Qy 509 PLMEKPLRPGSSGLPAPFSYFFVIRNAKVAAC 542
Db 556 PELKRPRLRAGRTLVIPVTMGFFVYKVNALAC 589

```

## RESULT 8

```

ID 08WMQ2 PRELIMINARY; PRT; 592 AA.
AC 08WMQ2;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Heparanase 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCB1_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Peasegue Salomae B.J.O.P.S.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN (12)
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Legoux P., Legoux R., O'Brien D., Salome M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ299719; CAC82491.1;
DR Genew; HGNC:18374; HPSB2.
DR InterPro; IPR005199; Glyco_hydro_79N.
DR Pfam; PF03662; Glyco_hydro_79N; I.
SQ SEQUENCE 592 AA; 66520 MW; 9478841FEACD5588 CRC64;

```

```

Query Match 40.3%; Score 1146.5; DB 4; Length 592;
Best Local Similarity 43.4%; Pred. No. 6.3e-83;
Matches 249; Conservative 82; Mismatches 190; Indels 53; Gaps 9;
20 PLGRLSPAL-----PRPA-----QAQDVVDLDFQTPRHLYVSS 55

```

```

Db 18 PPACIAPGALVYALALLHLHLSLSSOAGDRRLPVDBAAGLKEKTLILLDSTKRPVTVNEN 77
Qy 56 FLVSTIDANLADPRLILLGSPKRLTLAGLSPAYLRFGGTDTPLIF---DPKEST 111
Db 78 FLSTLQDPSIIHD-CWLDPLSSKRLVTLARGLSPALRFGKRTDPLQFONLRNPAKSRG 136
Qy 112 FEERSYMSQVUNODI-----CKYGSIPPDVEKRLLEMPYOEOL-LIREHYOK 158
Db 137 GGPDPYLLKNVEDDIIVRSVDVALDKQKCKIAQ-HPDVMLLEQREKAQOMHLVLEKEPSN 195
Qy 159 KFKNSTYSSVDVLYTFPANCGLDLIFGLNALRLTADLQNNSSNAQLLDYCSSKGYNI 218
Db 196 TYSNLIILARSLDKLYNFPADCGSLHLIFALNALRRPNNSNSSLILKYSAKKYNI 255
Qy 219 SWEIENEPNSFLKADIFINGSQJGEDIYQLHKLRLK-STFKNAKLYGPDVGQPRRTAK 277
Db 256 SWEIENEPNNRTYMGRAVNGSQJGKDYIQLKSLQPIRYSRSLYGPNGRBRKNVIA 315
Qy 278 MLKSFILKAGGEVIDSYMHYLLNGRTAREDFLNPVDLIFISSVOKVFOVVESTRPGK 337
Db 316 LLDGPMKAVGSTVDAVTMOHCYIDGRVVKVMDPLKTRLLDLSDOIIRKIQKVNTYTPGK 375
Qy 338 KWLGETSSAYGCGAPLSDTFAAGFMWLDKLSARMGIEVVRQVFFGAGNYHLVDEN 397
Db 376 KIMLEGVTTTSAAGTNNLSDSYAAGFLMNTLGMLANGIDIVIRHSFFDHGNYHLVDON 435
Qy 398 FDPPLDYLWLSLFLKLVGTGKVLMAVSOGSKR-----KLAVYLCTNTDNPRIYEG 448
Db 436 FNPPLDYLWLSLFLKLVGTGKVLMAVSOGSKR-----KLAVYLCTNTDNPRIYEG 495
Qy 449 DLTLYAINLHVNTKYLRLPYFPFSNKQYDKYLLRPLGPHGLSKSVQNLGLTKVWDDOTL 508
Db 496 SITPLTINLHRSRKKIKIAGTLRDKLWQYLLQPYQGEGLSKSVQNLGLTKVWDDOTL 555
Qy 509 PLMEKPLRPGSSGLPAPFSYFFVIRNAKVAAC 542
Db 556 PELKRPRLRAGRTLVIPVTMGFFVYKVNALAC 589

```

## RESULT 9

```

ID 08WMQ1 PRELIMINARY; PRT; 548 AA.
AC 08WMQ1;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Heparanase 3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCB1_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Peasegue Salomae B.J.O.P.S.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN (12)
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Legoux P., Legoux R., O'Brien D., Salome M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ299720; CAC82492.1;
DR InterPro; IPR005199; Glyco_hydro_79N.
DR Pfam; PF03662; Glyco_hydro_79N; I.
SQ SEQUENCE 548 AA; 61771 MW; B8986303FC7A60A CRC64;

```

```

Query Match 35.7%; Score 1015.5; DB 4; Length 548;
Best Local Similarity 41.8%; Pred. No. 1.7e-72;
Matches 224; Conservative 79; Mismatches 180; Indels 53; Gaps 9;
8 ALPPLMLLLGCPRLSPAL-----PRPA-----QAQDVVDLDF 43
6 AFPEMLSSNSRPPACIAPGALVYALALLHLHLSLSSOAGDRRLPVDBAAGLKEKTLILLD 65

```

QY 44 PROBLHVSFSLVSTIDANLATDPRFLILGSPKLTARGLSPAYLRFGTKTDFLI 103  
 DB 66 STKNPRTYVNEFLSLQDPSIHD-GMLDPLSSKRLVTLARGLSPAYLRFGTKTDFLI 124  
 QY 104 F----DPKKESTFEERSYWSQVNDI-----CKYGSIPDVEEKLRLWEPYQ 147  
 DB 125 FGNLNPASRSGPDYLLKNYEDIVASVDALDOKCKCIAO-HPDVMELOREKAAQ 183  
 QY 148 EOL-LIREHYOKKFKKSTYSRSSVDLYTFANCSGLDIFGLMALLRTADLWNSNAOL 206  
 DB 184 MELVLLKEQPSYTSYLITARSGLDLYNSADCSGLHLFALNALRRNNNSWNSALS 243  
 QY 207 LIDYSSKSGYNI-SWELGNEPNSFLKADIFINGSQGEDEYIOLHKLK-RSTFNKAKYQ 265  
 DB 244 LKTSASKRYNISWELGNEPNSFLKADIFINGSQGEDEYIOLHKLK-RSTFNKAKYQ 303  
 QY 266 PDVGPRRRTAKMLKSLKAGSEVIDSVTMHYYLNGRTAREDEFLNPDVLDIFISSVOK 325  
 DB 304 PIGPRKRVIALLOGFVKVAGSTVDAYWOCYIDGRVVKWMDPLKTRLDLTLSDQIRK 363  
 QY 326 VPOVSESTRPGKRWLGSTSSAYGGAFLSDTPAAGFWMDKLGSAENGIEVMKQVY 385  
 DB 364 IOKVNTYTPGKKIMLEGVVTTSAGTNNLSDSYAAGFLMLNTLGLANOGIDVYIRHSF 423  
 QY 366 PCAGNYHLVDENFDPLPDYMLSLPFKLVGTKYLMASVOGSKR-----KLRYLH 436  
 DB 424 PFGHNLHLDVONFNPDPYMLSLTLKRLGPKYLAHVAGLQKPRPGVYIDKRIYH 483  
 QY 437 CTNTDNPYKEDDLTYALNLHNTYKRLPYFPNSKQVDTYLRPLGPHGLSKS 492  
 DB 484 CTNNHNNYVRSITLFIINLHRSKKIKLAGTLRKLVHQYLLQYGOGLSKSKT 539

## RESULT 10

Q9HB38 PRELIMINARY; PRT; 534 AA.  
 ID AC Q9HB38;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DB Hepatanae-like protein HPA2b.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCBITaxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20483645; PubMed=11027606;  
 RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,  
 RA Hircok M., Patel S., Barry E., Stuberfield C., Terrett J., Page M.,  
 RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian  
 RT Hepatanae Family Member."  
 RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).  
 DR EMBL; AF282886; AAC23422.1;  
 DR InterPro: IPR005199; Glyco\_hydro\_79n;  
 DR Pfam: PF03662; Glyco\_hydro\_79n;  
 SQ SEQUENCE 534 AA; 60063 MW; C3DSEB900CB38C4 CRC64;

Query Match 33.0%; Score 936.5; DB 4; Length 534;  
 Best Local Similarity 37.8%; Pred. No. 3.3e-66;  
 Matches 217; Conservative 78; Mismatches 168; Indels 111; Gaps 12;

QY 20 PLGSPSPAL-----PRPA-----QAQDVVDLDFTOEPRLHVS 55  
 DB 18 PACIAPALVIALHLHLSLSSQAGRRPLPVDRAAGLKEKTLILDVSTKMPVTVNEN 77  
 QY 56 FLSTVDANLATDPRFLILGSPKLTARGLSPAYLRFGTKTDFLI-----DPKKEST 111  
 DB 78 FLSTQDPSIHD-GMLDPLSSKRLVTLARGLSPAYLRFGTKTDFLIQONTLNPASRSG 136  
 QY 112 FEERSYWSQVNDI-----CKYGSIPDVEEKLRLWEPYQOL-LIREHYOK 158

DB 137 GPGPDYLLKNYEDIVASVDALDOKCKCIAO-HPDVMELOREKAAQMLVLLKEQF-- 193  
 QY 159 KFKSTYSRSSVDLYTFANCSGLDIFGLMALLRTADLWNSNAOLLDYSSKGYNI 218  
 DB 194 ---SNTYS-----NLTL----- 202  
 QY 219 SWELGNEPNSFLKADIFINGSQGEDEYIOLHKLK-RSTFNKAKYGPDPGPRRTAK 277  
 DB 203 ---TEPNYRTMNGAIVNSQGLKQYIOLKSLLOPFIYRSASLVGPNIGRRKQVIA 257  
 QY 278 MLKSPFLKAGSEVIDSVTMHYYLNGRTAREDEFLNPDVLDIFISSVOKVPOVSESTRPG 337  
 DB 258 LIDGFMKVAAGSTVDAYWOCYIDGRVVKWMDPLKTRLDLTLSDQIRKIOCVNTYTPGK 317  
 QY 338 KWLIGETSSAYGGAFLSDTPAAGFWMDKLGSAENGIEVMKQVYFGAGNYLVDEN 397  
 DB 318 KIMLGVTTSAGTNNLSDSYAAGFLMLNTLGLANOGIDVYIRHSFDFGYNHLDON 377  
 QY 398 FDPPLDYMLSLPFKLVGTKYLMASVOGSKR-----KLRYLHCTNTDNPYK 448  
 DB 378 FNPPLDYMLSLTLKRLGPKYLAHVAGLQKPRPGVYIDKRIYHACTNNHNNYVRS 437  
 QY 449 DLTVALNLHNTYKRLPYFPNSKQVDTYLRPLGPHGLSKSYOLNGLTLKAYDDOTL 508  
 DB 438 SITLFIINLHRSKKIKLAGTLRKLVHQYLLQYGOGLSKSKSYOLNGLTLKAYDDOTL 497  
 QY 509 PLMEKPLRPGSSGLPAPFSFYIRAKVLAAC 542  
 DB 498 PELKPRPLRAGRTLVIPVTWGFVAKVNALAC 531

## RESULT 11

Q9HB39 PRELIMINARY; PRT; 480 AA.  
 ID AC Q9HB39;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DB Hepatanae-like protein HPA2a.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCBITaxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20483645; PubMed=11027606;  
 RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,  
 RA Hircok M., Patel S., Barry E., Stuberfield C., Terrett J., Page M.,  
 RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian  
 RT Hepatanae Family Member."  
 RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).  
 DR EMBL; AF282885; AAC23422.1;  
 DR InterPro: IPR005199; Glyco\_hydro\_79n;  
 DR Pfam: PF03662; Glyco\_hydro\_79n;  
 SQ SEQUENCE 480 AA; 53900 MW; F75F89F67AC1FF83 CRC64;

Query Match 31.6%; Score 897.5; DB 4; Length 480;  
 Best Local Similarity 36.0%; Pred. No. 3.7e-63;  
 Matches 202; Conservative 74; Mismatches 146; Indels 139; Gaps 9;

QY 20 PLGSPSPAL-----PRPA-----QAQDVVDLDFTOEPRLHVS 55  
 DB 18 PACIAPALVIALHLHLSLSSQAGRRPLPVDRAAGLKEKTLILDVSTKMPVTVNEN 77  
 QY 56 FLSTVDANLATDPRFLILGSPKLTARGLSPAYLRFGTKTDFLI-----DPKKEST 111  
 DB 78 FLSTQDPSIHD-GMLDPLSSKRLVTLARGLSPAYLRFGTKTDFLIQONTLNPASRSG 135  
 QY 112 FEERSYWSQVNDI-----CKYGSIPDVEEKLRLWEPYQOL-LIREHYOK 171  
 DB 136 -----GPGPD-----YLRKYE----- 148  
 QY 172 VLYTFANCSGLDIFGLMALLRTADLWNSNAOLLDYSSKGYNISWELGNEPNSFLK 231

Db 42 OBDIKT.

3

•

QY 135 DVEBKRLNMPYQBOILLREHYOKKFKNS---TYSRSSV---DVLTYFANCGLDLIF 186  
 Db -----PFOKM-----NSGLFPGSKCGLMKRMDELNSFLTAGAVTF 132  
 QY 95 -----PFOKM-----NSGLFPGSKCGLMKRMDELNSFLTAGAVTF 132  
 Db 187 GINALLRTADLQ-----MNSSNAQLLLDYCSSKGYNI-SWEIGNEPNSFLKADIFIN 238  
 QY 133 GINALRGHRLRKAGWAGAMDHINTODFLNNTVSKGYVIDSWEFGNELSG--SGVGASVS 190  
 Db 239 GSQLEDYIQLHLKLRSTFKNAKLYGPDVGP-----RRKTAOKLSFLKAGGEVIDSV 293  
 QY 191 AELYGKDLVLKQVINK-VYKNSWLHKPILVAPGFGYEQWYTKLLEI---SGPSVADV 246  
 Db 294 TWHYYLNGRT--ATREDFLNPVDLIFISSVQVF---QVESTRPCKVWLGSETSA 347  
 QY 247 THHIVNLGSGNDPALVKKIMDP-----YLSQVSKTFKDVNQTIOEHGPMASPMVGESGA 302  
 Db 348 YGGGAPLLSDPFAAGFMWLDLGLSABMGIEVWQVFFGAGNYHLVDE-NEDPLPDYWL 406  
 QY 303 YNSGGRHVSDTFIDSFYWLDDQLGMSARHNTKYVCROTLVG--GFYGLLEKGTFFVNPDDYS 361  
 Db 407 SLLEKLVGTQVTLVMAVSQSKRKLRYVLIHCTNTDNPYKEGDLTLVAINLHNTKYL-- 464  
 QY 362 ALLMRLMKGKVLAVQTDGP--QLRVYAHCSK-----GRAGVTLLILNLSQSDFTVS 413  
 Db 465 -----RLPYPS--NKQVDKYLRLP--LGPFG--LLSKSVOL 495  
 QY 414 VSNGINVVLNABSRKKSLDLTLKRPFSWIGSKASDGYLNEEYHLTPENGVLRSKTVWL 473  
 Db 496 NGTLTKWDDQTLPLMEKPLRP-GSSGLPAFSYFPVIRNAKVAAC 542  
 QY 474 NKSILKPTATGDIPLS-LPVLRSVNSPLNVLPLSMFSFVLNPDASAC 520

## RESULT 14

Q9PFI0 PRELIMINARY; PRT; 543 AA.  
 AC Q9PFI0;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Similarity to heparanase.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=97471969; PubMed=9330910;  
 RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asanizu E., Fukami M.,  
 Miyajima N., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
 features of the 1.6 Mb regions covered by twenty physically assigned  
 RT clones."  
 RT DNA Reg. 4:215-230(1997).  
 DR EMBL; AB005249; BAB09847.1;  
 DR InterPro; IPR005199; Glyco\_hydro\_79N.  
 DR InterPro; IPR001254; Ser\_protease\_tty.  
 DR Pfam; PF03662; Glyco\_hydro\_79n; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SRR; 1.  
 SQ SEQUENCE 543 AA; 60250 MW; 0FA2248948282FF6 CRC64;

Query Match 14.6%; Score 416; DB 10; Length 543;

Best Local Similarity 29.2%; Pred. No. 1,4e-24;

Matches 154; Conservative 68; Mismatches 184; Indels 122; Gaps 24;

75 LGSPLKRTIARGLSPAYLRFGGTKTDFLFDPKESTFEERSYWGQVNOIDCKYCSIPP 134

Db 77 LRPPLTKAIKAKPARIKIGSLQDOVIYDGNLKT-----PCR----- 116

QY 135 DVEBKRLNMPYQBOILLREHYOKKFKNS---TYSRSSV---DVLTYFANCGLDLIF 186

Db 117 -----PFOKM-----NSGLFPGSKCGLMKRMDELNSFLTAGAVTF 154  
 QY 187 GINALLRTADLQ-----MNSSNAQLLLDYCSSKGYNI-SWEIGNEPNSFLKADIFIN 238  
 Db 155 GINALRGHRLRKAGWAGAMDHINTODFLNNTVSKGYVIDSWEFGNELSG--SGVGASVS 212  
 QY 239 GSQLEDYIQLHLKLRSTFKNAKLYGPDVGP-----RRKTAOKLSFLKAGGEVIDSV 293  
 Db 213 AELYGKDLVLKQVINK-VYKNSWLHKPILVAPGFGYEQWYTKLLEI---SGPSVADV 268  
 QY 294 TWHYYLNGRT--ATREDFLNPVDLIFISSVQVF---QVESTRPCKVWLGSETSA 347  
 Db 269 THHIVNLGSGNDPALVKKIMDP-----YLSQVSKTFKDVNQTIOEHGPMASPMVGESGA 324  
 QY 348 YGGGAPLLSDPFAAGFMWLDLGLSABMGIEVWQVFFGAGNYHLVDE-NEDPLPDYWL 406  
 Db 335 YNSGGRHVSDTFIDSFYWLDDQLGMSARHNTKYVCROTLVG--GFYGLLEKGTFFVNPDDYS 383  
 QY 407 SLLEKLVGTQVTLVMAVSQSKRKLRYVLIHCTNTDNPYKEGDLTLVAINLHNTKYL-- 464  
 Db 364 ALLMRLMKGKVLAVQTDGP--QLRVYAHCSK-----GRAGVTLLILNLSQSDFTVS 435  
 QY 465 -----RLPYPS--NKQVDKYLRLP--LGPFG--LLSKSVOL 495  
 Db 436 VSNGINVVLNABSRKKSLDLTLKRPFSWIGSKASDGYLNEEYHLTPENGVLRSKTVWL 495  
 QY 436 NGTLTKWDDQTLPLMEKPLRP-GSSGLPAFSYFPVIRNAKVAAC 542  
 Db 436 NKSILKPTATGDIPLS-LPVLRSVNSPLNVLPLSMFSFVLNPDASAC 542

## RESULT 15

Q9H615 PRELIMINARY; PRT; 544 AA.  
 AC Q9H615;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE OSJUNBA0035103.11 protein.  
 GN OSJUNBA0035103.11.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzae; Oryza.  
 OC NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, BAC  
 RT clone:OSJUNBA0035103."  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP003019; BAC22238.1;  
 SQ SEQUENCE 544 AA; 58723 MW; ECD8695F0E22A269 CRC64;

Query Match 13.8%; Score 392.5; DB 10; Length 544;

Best Local Similarity 26.8%; Pred. No. 1e-22;

Matches 144; Conservative 58; Mismatches 179; Indels 157; Gaps 19;

75 LGSPLKRTIARGLSPAYLRFGGTKTDFLFD-----PKESTFEERSYWGQVNOID 125

Db 93 LSNKILNAINAFSPFLKRLIGSLQDRLVYGDDGCGCAFFVKNVTSMPFGPTQ----- 146

QY 126 ICKYSIIPDVBEKRLNMPYQBOILLREHYOKKFKNSTYSSSDVLYTFANCGLDLIF 185

Db 147 -----GCLPLH-----RW-----DELNAFFQSGARIV 169

QY 186 FGLNALLRTADLQ-----MNSSNAQLLLDYCSSKGYNI-SWEIGNEPNSFLKADIFIN 237

Db 170 FGLNALLNRPVLPDQSGMGPMDYTNAAALIRYTSKGIKIGWELGN-----L 218

QY 238 NSQLEDYIQLHLKLRSTFKNAKLYGPDV-----QPRRTAKMLKSLFLKA 285



06:21:55 2003

```

Db      219 SGSVC-----TKVGADQYADVIALKSLVDTIYQGNPSKPLVLA PGFPDA 265
QY      286 G--GEVID-----SVTWHHYYLNG--RTATREDFLNPDVLDIFISSVQKVFQVSEST 333
Db      266 GWFTBEVIVKTRPNLLNVVTHHLYNLGPGVDTHLIEKILNPSTLDGKVSSTFSLQGLKSA 325
QY      334 RPKKKVWLGETSAYVGGAPLSDTFAAGFMMLDKLGSABMGIEVVMRQVFFGAGNYHL 393
Db      326 GTSAAVAVGESSGAYNSGRHLVTDSEVFSFWMFLDQJGMSAKYDTKSYCRQSLIG-GNYGL 384
QY      394 VD-ENFDPLPDYWLSLFLFKLVGTVKVLMSVOGSKRRKLRVYLHCTNTONPRYKBDLTL 452
Db      385 LNKETQPNPDYYSALLMHRMLNGTKVLSATFNCT--NMIRTYAHCAK-DSP-----GITL 436
QY      453 YAINL-----HNTKYLRLPYPPSNQVQDKYLLRPIG 484
Db      437 LILNLSGNTTSQVSVTSEGAHANTVKKHSRKTTHLAAGSRREBYHLJAKD----- 486
QY      485 PHGLSKSVQNLGLTLKMDVDDTLPLMEKRLRPGSSIGLPAPSYSPFYIRNAKVAAC 542
Db      487 -GSLQSQVMLNGRALVADENGELPRLEPVKVDAAOPIAVAPYSIVFAHIHNFPA PAC 543

```

Search completed: October 22, 2003, 20:26:50  
 Job time : 110 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2003, 20:19:45 ; Search time 86 Seconds  
(without alignments)  
1002.192 Million cell updates/sec

Title: US-09-759-207-2

Perfect score: 1 MLRSKRALPPPLMLLGP.....LPARISFPIRNKVAACI 543

Sequence: 1 MLRSKRALPPPLMLLGP.....LPARISFPIRNKVAACI 543

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

A\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*

2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*

3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*

4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*

5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*

6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*

7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*

8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*

9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*

10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*

11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*

12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*

13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*

14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*

15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*

16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*

17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*

18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*

19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*

20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*

21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*

22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2842	100.0	543	20: AAY02345	A human heparanase
2	2842	100.0	543	21: AAB08849	Amino acid sequenc
3	2842	100.0	543	21: AAY57590	Human heparanase.
4	2842	100.0	543	21: AAY52990	Human heparanase p
5	2842	100.0	543	22: AAY97635	Human heparanase p
6	2842	100.0	543	23: ABB07813	Human heparanase s
7	2842	100.0	592	20: AAY02346	A human heparanase
8	2842	100.0	592	21: AAB08850	Amino acid sequenc
9	2838	99.9	543	20: AAY17082	Human heparanase e

10	2838	99.9	543	22: AAB86206	Human heparanase 1
11	2838	99.9	588	20: AAY30124	A human protein w/
12	2826	99.4	543	22: AAB88361	Human membrane or
13	2817	99.1	545	24: ABB56822	Human heparanase p
14	2764	97.3	530	20: AAY34173	Human pre-prohepar
15	2737	96.3	532	20: AAY17083	Seg ID No: 15 of W
16	2671.5	94.1	527	23: ABB07815	Chicken signal pep
17	2146	75.5	535	21: AAB08851	A murine heparanas
18	2146	75.5	535	23: ABB07811	Mouse heparanase s
19	2123	74.7	536	23: ABB07812	Rat heparanase seq
20	1645.5	57.9	523	23: ABB07814	Chicken heparanase
21	1614	56.8	380	20: AAY17085	Mouse heparanase e
22	1602	56.4	380	20: AAY17084	Human heparanase e
23	1154.5	40.6	592	22: AAY07424	Human heparanase-1
24	1154.5	40.6	592	22: AAY97632	Human heparanase-2
25	1148.5	40.4	592	22: AAB81062	Human heparanase-2
26	1147.5	40.4	592	22: AAB85215	Heparanase-like pr
27	1142.5	40.2	582	23: AAB18326	Human heparanase-2
28	1112.5	39.1	538	22: AAY97633	Human heparanase,
29	1106.5	38.9	528	23: AAB18327	Human heparanase-2
30	936.5	33.0	534	22: AAB85216	Heparanase-like pr
31	936.5	33.0	534	23: AAB69310	Human polypeptide
32	936.5	33.0	534	23: AAM50337	Human prepro-hepar
33	927.5	32.6	492	22: AAB84664	Amino acid sequenc
34	897.5	31.6	480	22: AAY07418	Novel human extrac
35	897.5	31.6	480	22: AAB85217	Heparanase-like pr
36	887.5	31.6	480	22: AAY97634	Human heparanase,
37	892.5	31.4	470	23: AAB18328	Human heparanase-2
38	891.5	31.4	439	23: AAY07423	Human heparanase-1
39	788	27.7	331	23: AAM50383	Human heparanase I
40	663	23.3	488	22: AAB31469	Amino acid sequenc
41	645	22.7	488	22: AAB31470	Amino acid sequenc
42	642	22.6	488	22: AAB31472	Amino acid sequenc
43	622	21.9	488	22: AAB31471	Amino acid sequenc
44	528.5	18.6	214	22: AAM99905	Human excretory re
45	528.5	18.6	214	22: AAM43704	Human bladder anti

## ALIGNMENTS

RESULT 1

AA02345

ID AAY02345 standard; Protein: 543 AA.

XX AAY02345;

XX

XX 09-JUL-1999 (first entry)

XX

XX A human heparanase protein.

DE

XX Heparanase; hp; modulator; heparin-binding growth factor;

XX cellular response; cytokine; cell interaction; plasma lipoprotein;

KW cellular susceptibility; infection; distention;

KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;

KW atherosclerosis; inflammation; neurodegenerative disease; neuritis;

KW plasma heparin; microtubulastasis; autoimmune lesion; renal failure.

XX

XX Homo sapiens.

OS

XX

XX PD WO911798-A1.

XX

XX 11-MAR-1999.

XX

XX PF 31-AUG-1998; 98WO-US17954.

XX

XX 02-JUL-1998; 98US-0109386.

XX

XX 02-SEP-1997; 97US-0922170.

XX

XX (FRIE/) FRIEDMAN M M.

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

PA (INST-) INSIGHT STRATEGY & MARKETING LTD.

XX

PI Feinstein E, Pecker I, Vlodavsky I;  
 XX WPI; 1999-302255/25.  
 DR N-PSDB; AAX35648.  
 XX  
 PT New human polynucleotide useful for treating angiogenesis,  
 XX restenosis, and inflammation  
 PS  
 PS Claim 6; Fig 1; 63pp; English.  
 XX The specification describes a polypeptide having heparanase (hp)  
 CC activity. The recombinant protein is used as a modulator of  
 CC heparin-binding growth factors, cellular responses to heparin-binding  
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
 CC cellular susceptibility to viral, protozoal and bacterial infections  
 CC or disturbance of neurodegenerative plaques. Heparanase may be  
 CC useful for conditions such as wound healing, angiogenesis, restenosis,  
 CC atherosclerosis, inflammation, neurodegenerative diseases, and viral  
 CC infections. Mammalian heparanase can be used to neutralize plasma  
 CC heparin, and anti-heparanase antibodies may be applied for  
 CC immunodetection and diagnosis of micrometastases, autoimmune lesions,  
 CC and renal failure in biopsy specimens, plasma samples, and body fluids.  
 CC The present sequence represents human heparanase.  
 XX  
 XX Sequence 543 AA;  
 SQ  
 Query Match 100.0%; Score 2842; DB 20; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-274;  
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MLRSKPLPPPLMLLLGLPGLSPGALPRPAQADVDLDFPQEPHLVSPSFLSVT 60  
 DB 1 MLRSKPLPPPLMLLLGLPGLSPGALPRPAQADVDLDFPQEPHLVSPSFLSVT 60  
 QY 61 IDANLATDPRLILGSPKRLTARGLSPAYLRFGGTDFLFDPKKSTFEERSYWS 120  
 DB 61 IDANLATDPRLILGSPKRLTARGLSPAYLRFGGTDFLFDPKKSTFEERSYWS 120  
 QY 121 QVNODICKYGISPPVBEKRLLEMPYQQLREHYQKKFKNSYSSVDVLTTPANC 180  
 DB 121 QVNODICKYGISPPVBEKRLLEMPYQQLREHYQKKFKNSYSSVDVLTTPANC 180  
 QY 121 QVNODICKYGISPPVBEKRLLEMPYQQLREHYQKKFKNSYSSVDVLTTPANC 180  
 DB 121 QVNODICKYGISPPVBEKRLLEMPYQQLREHYQKKFKNSYSSVDVLTTPANC 180  
 QY 181 GLDLIFGALNLRADLQNMSSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFNGS 240  
 DB 181 GLDLIFGALNLRADLQNMSSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFNGS 240  
 QY 241 QLGEDYIOLHLKRLKSTFKNAKLYGPDVGOPPRKAKMLKSFLLKAGEVIDSVTHHYLL 300  
 DB 241 QLGEDYIOLHLKRLKSTFKNAKLYGPDVGOPPRKAKMLKSFLLKAGEVIDSVTHHYLL 300  
 QY 301 NGRTRAREDPINPDVLDIFISSVQKRVFVVESTPRGKVMIGETSSAYGGAFLSDTFA 360  
 DB 301 NGRTRAREDPINPDVLDIFISSVQKRVFVVESTPRGKVMIGETSSAYGGAFLSDTFA 360  
 QY 361 AGFWMLDLGLSABWGVVWROVFEFGAGNHLVDENDDLPDWLSTLLFKKVGSTKYL 420  
 DB 361 AGFWMLDLGLSABWGVVWROVFEFGAGNHLVDENDDLPDWLSTLLFKKVGSTKYL 420  
 QY 421 ASVQSKRKRLRVYLACTNTDPRRYEGDLTLVAINLHNTKYRLPYPFNSKQVDKYL 480  
 DB 421 ASVQSKRKRLRVYLACTNTDPRRYEGDLTLVAINLHNTKYRLPYPFNSKQVDKYL 480  
 QY 481 RPLGPHGLLSKSVQVINGLTAKVDDQTLPLMEKPLRGSSIGLPAFYSFVIRNAKVA 540  
 DB 481 RPLGPHGLLSKSVQVINGLTAKVDDQTLPLMEKPLRGSSIGLPAFYSFVIRNAKVA 540  
 QY 541 ACT 543  
 DB 541 ACT 543

ID AAB08849 standard; Protein; 543 AA.  
 XX AAB08849;  
 AC  
 XX 15-JAN-2001 (first entry)  
 DT  
 XX  
 DE Amino acid sequence of a human heparanase polypeptide.  
 XX  
 XX Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;  
 KM heparin-binding growth factor; cytokine; neurodegenerative plaque;  
 KM wound healing; infection; burn; angiogenesis; restenosis;  
 KM atherosclerosis; inflammation; neurodegenerative disease;  
 KM Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.  
 XX  
 OS Homo sapiens.  
 XX  
 XX MO200052178-A1.  
 XX  
 XX 08-SEP-2000.  
 PD  
 XX  
 XX 14-FEB-2000; 2000MO-US03542.  
 PF  
 XX  
 XX 01-MAR-1999; 99US-0258892.  
 PR  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA (FRIE/) FRIEDMAN M M.  
 XX  
 XX Pecker I, Vlodavsky I, Feinstein E;  
 XX  
 XX WPI; 2000-579289/54.  
 DR N-PSDB; AAX75051.  
 XX  
 XX New polynucleotides encoding a polypeptide having heparanase activity,  
 PT useful in wound healing and in gene therapy, particularly in treating  
 PT tumour, inflammation, autoimmunity, neurodegenerative diseases  
 XX  
 PS Claim 22; Fig 1; 152pp; English.  
 XX  
 XX The present sequence represents a human protein with heparanase catalytic  
 CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,  
 CC particularly in treating tumour, inflammation or autoimmunity.  
 CC Particularly, the polynucleotide is useful in modulating the  
 CC bioavailability of heparin-binding growth factors, cellular responses  
 CC to heparin-binding growth factors (e.g. bFGF) and cytokines  
 CC (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins,  
 CC cellular susceptibility to certain viral and some bacterial and protozoa  
 CC infections, or disintegration of neurodegenerative plaques. The  
 CC polynucleotide is also useful in wound healing (e.g. thermal, chemical  
 CC or radiation burns), and in the treatment of angiogenesis, restenosis,  
 CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-  
 CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,  
 CC bacterial or protozoa infections.  
 XX  
 XX Sequence 543 AA;  
 SQ  
 Query Match 100.0%; Score 2842; DB 21; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-274;  
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MLRSKPLPPPLMLLLGLPGLSPGALPRPAQADVDLDFPQEPHLVSPSFLSVT 60  
 DB 1 MLRSKPLPPPLMLLLGLPGLSPGALPRPAQADVDLDFPQEPHLVSPSFLSVT 60  
 QY 61 IDANLATDPRLILGSPKRLTARGLSPAYLRFGGTDFLFDPKKSTFEERSYWS 120  
 DB 61 IDANLATDPRLILGSPKRLTARGLSPAYLRFGGTDFLFDPKKSTFEERSYWS 120  
 QY 121 QVNODICKYGISPPVBEKRLLEMPYQQLREHYQKKFKNSYSSVDVLTTPANC 180  
 DB 121 QVNODICKYGISPPVBEKRLLEMPYQQLREHYQKKFKNSYSSVDVLTTPANC 180  
 QY 181 GLDLIFGALNLRADLQNMSSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFNGS 240

DB 181 GIDLIFGIALNLRTRADLOMNSNAQLLDYCSSKGNINISWEIGNEPNSFLKKAIDIFINGS 240  
 QY 241 QLEGEDYIOLHLKLRKSTFKNAKLYGPDVGOPRRKTAAMKLSFLKAGGEVIDSVTHHHYLL 300  
 DB 241 QLEGEDYIOLHLKLRKSTFKNAKLYGPDVGOPRRKTAAMKLSFLKAGGEVIDSVTHHHYLL 300  
 QY 301 NGRTATREDPLNPVDLDFIISVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360  
 DB 301 NGRTATREDPLNPVDLDFIISVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360  
 QY 361 AGFWMULDGLGSARNGIEVVMROVFFGAGNYHLVDENPDLPDYLSTLFFKLVGTYKLM 420  
 DB 361 AGFWMULDGLGSARNGIEVVMROVFFGAGNYHLVDENPDLPDYLSTLFFKLVGTYKLM 420  
 QY 421 ASVQSSKRRKRLRVYLHCTNTDNPVRYEGDLTYAINLHNVTKYLRPLPFSNKQVDKTYL 480  
 DB 421 ASVQSSKRRKRLRVYLHCTNTDNPVRYEGDLTYAINLHNVTKYLRPLPFSNKQVDKTYL 480  
 QY 481 RPLGPHGLSKSVQVQNGITLTKAVDDOTLPPLMEKPLRGSSIGLPAFYSFFVIRNAKYA 540  
 DB 481 RPLGPHGLSKSVQVQNGITLTKAVDDOTLPPLMEKPLRGSSIGLPAFYSFFVIRNAKYA 540  
 QY 541 ACT 543  
 DB 541 ACT 543

RESULT 3  
 AAY57590  
 ID AAY57590 standard; Protein; 543 AA.

DT 02-MAR-2000 (first entry)

DE Human heparanase.

XX Human; heparanase; hpa; genetic modification; expression; anticancer;  
 KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumor;  
 KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;  
 KW heparin sulphate; heparin-binding growth factor; tumour angiogenesis;  
 KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;  
 KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;  
 KW micrometastasis; autoimmune lesion; kidney failure.

OS Homo sapiens.

XX W09557244-A1.

XX 11-NOV-1999.

XX 29-APR-1999; 99WC-US09256.

XX 01-MAY-1998; 98US-0071618.

XX 02-MAR-1999; 99US-0260038.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX (FRIE-) FRIEDMAN M M.

XX Ben-Artzi H, Ayal-HersHKovitz M, Yacoby-Zeevi O, Pecker I, Peleg Y,

XX Shlom Y;

XX WPI: 2000-062144/05.

XX N-PSDB; AAC39195.

XX Engineered cells that express recombinant heparanase, useful

XX therapeutically, e.g. for treating angiogenesis and to screen for

XX specific inhibitors, potential anticancer agents

XX Claim 3; Page 107-109; 118pp; English.

XX The present invention describes genetically modified cells (A) containing

CC a polynucleotide (I) that encodes a polypeptide with heparanase activity,  
 CC and expresses recombinant heparanase (II). Heparanase cleaves heparan  
 CC sulphate (HS) at specific intrachain sites, resulting in release of  
 CC heparin-binding growth factors, enzymes and proteins that are sequestered  
 CC by HS in basement membranes, extracellular matrix or cell surfaces. It  
 CC may also be implicated in tumour angiogenesis and metastasis. (II) is  
 CC potentially useful in wound healing and for treating angiogenesis,  
 CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral  
 CC infection and cystic fibrosis. It can also be used to neutralise heparin  
 CC (an alternative to protamine) and to screen for specific inhibitors  
 CC (potentially useful for treating cancer and metastasis). Antibodies  
 CC raised against (II) are used for immunodetection and diagnosis of  
 CC micrometastases, autoimmune lesions and kidney failure. (A) provide (II)  
 CC in large quantities, in a form that is homogeneously processed and  
 CC activated/neutralised by a dedicated protease. The present sequence  
 CC represents human heparanase.

Sequence 543 AA;

Query Match 100.0%; Score 2842; DB 21; Length 543;

Best Local Similarity 100.0%; Pred. No. 9.5e-274;

Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKPLPPPLMLLGLGPIGSPGALPRPAQADVDLDFPQEPHLVSPSFLSVT 60

DB 1 MLRSKPLPPPLMLLGLGPIGSPGALPRPAQADVDLDFPQEPHLVSPSFLSVT 60

QY 61 IDANLATDPRFLILGSPKRLTARGSPAYLRFQGTDFLIPDKKSTFEERSYQMS 120

DB 61 IDANLATDPRFLILGSPKRLTARGSPAYLRFQGTDFLIPDKKSTFEERSYQMS 120

QY 121 QVNODICKYGSIPDVEBKRLLEWYQEQLLREHYQKKFNSTYSRSSVDVLYTFANCS 180

DB 121 QVNODICKYGSIPDVEBKRLLEWYQEQLLREHYQKKFNSTYSRSSVDVLYTFANCS 180

QY 181 GIDLIFGIALNLRTRADLOMNSNAQLLDYCSSKGNINISWEIGNEPNSFLKKAIDIFINGS 240

DB 181 GIDLIFGIALNLRTRADLOMNSNAQLLDYCSSKGNINISWEIGNEPNSFLKKAIDIFINGS 240

QY 241 QLEGEDYIOLHLKLRKSTFKNAKLYGPDVGOPRRKTAAMKLSFLKAGGEVIDSVTHHHYLL 300

DB 241 QLEGEDYIOLHLKLRKSTFKNAKLYGPDVGOPRRKTAAMKLSFLKAGGEVIDSVTHHHYLL 300

QY 301 NGRTATREDPLNPVDLDFIISVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360

DB 301 NGRTATREDPLNPVDLDFIISVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360

QY 361 AGFWMULDGLGSARNGIEVVMROVFFGAGNYHLVDENPDLPDYLSTLFFKLVGTYKLM 420

DB 361 AGFWMULDGLGSARNGIEVVMROVFFGAGNYHLVDENPDLPDYLSTLFFKLVGTYKLM 420

QY 421 ASVQSSKRRKRLRVYLHCTNTDNPVRYEGDLTYAINLHNVTKYLRPLPFSNKQVDKTYL 480

DB 421 ASVQSSKRRKRLRVYLHCTNTDNPVRYEGDLTYAINLHNVTKYLRPLPFSNKQVDKTYL 480

QY 481 RPLGPHGLSKSVQVQNGITLTKAVDDOTLPPLMEKPLRGSSIGLPAFYSFFVIRNAKYA 540

DB 481 RPLGPHGLSKSVQVQNGITLTKAVDDOTLPPLMEKPLRGSSIGLPAFYSFFVIRNAKYA 540

QY 541 ACT 543

DB 541 ACT 543

RESULT 4

ID AAY52990 standard; Protein; 543 AA.

XX AAY52990;

XX 21-FEB-2000 (first entry)

XX Human heparanase protein sequence.

XX Human: heparanase; hpa; diagnosis; therapy; tumour; cytostatic;  
 KW anti-diabetic; immunomodulatory; anti-inflammatory; nephrotoxic;  
 KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;  
 KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;  
 KW inflammation; haemorrhagic nephritis; nephrotic syndrome;  
 KW autoimmune disease; anticancer; kidney disease.  
 XX Homo sapiens.  
 OS  
 XX MO9957153-A1.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PD 29-APR-1999; 99MO-US09255.  
 XX  
 XX 01-MAY-1998; 98US-0071739.  
 XX  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA (FRIE/) FRIEDMAN M M.  
 XX  
 XX Pecker I, Vlodavsky I, Friedman Y, Perets T;  
 PI  
 XX WPI: 2000-052934/04.  
 XX  
 DR N-PDB; AA233290.  
 XX  
 PT Heparanase-specific molecular probes useful for diagnosis and  
 PT treatment, e.g. of tumours, and for targeted drug delivery  
 XX  
 PS Example; Page 81-82; 90pp; English.  
 XX  
 CC The present invention describes heparanase-specific molecular probes,  
 CC useful for methods of detecting heparanase in situ. The probes and  
 CC anti-heparanase antibodies are used to detect or quantify the expression  
 CC of heparanase, for diagnosis and monitoring of diseases (especially  
 CC metastasis), for treatment of heparanase-associated diseases (e.g.  
 CC tumours, (adenocarcinoma, squamous cell carcinoma, teratocarcinoma,  
 CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its  
 CC metastases) derived from liver, prostate, bladder, breast, ovary,  
 CC cervix, colon, skin, intestine, stomach, uterus and pancreas, kidney  
 CC disease, diabetes and inflammation, haemorrhagic nephritis, nephrotic  
 CC syndrome, sepsis and inflammatory or autoimmune disease), for targeted  
 CC drug delivery (e.g. of anticancer agents) and as research reagents.  
 CC The present sequence represents human heparanase, which is used in the  
 CC exemplification of the present invention.  
 CC  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 2842; DB 21; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-274;  
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MLRSKRALPPMLLLGLPLSPGALPRPAQADVLDLFTQEPHLVSPSLSVT 60  
 DB 1 MLRSKRALPPMLLLGLPLSPGALPRPAQADVLDLFTQEPHLVSPSLSVT 60  
 QY 61 IDANLATDPPFLILGSPKRTIARGLSPAYLRFGTGTFLIPDKKSTFEERSYWG 120  
 DB 61 IDANLATDPPFLILGSPKRTIARGLSPAYLRFGTGTFLIPDKKSTFEERSYWG 120  
 QY 121 QVNVODICKYSSIPDVBEKRLRLEMPYQOLLEHVKRKNSTYSSVDVLYTPANCS 180  
 DB 121 QVNVODICKYSSIPDVBEKRLRLEMPYQOLLEHVKRKNSTYSSVDVLYTPANCS 180  
 QY 181 GDPLIFGIALNLRTADIQMNSSNAQLLDYCSSKGYNISWELNEPNPSFKKADIFINS 240  
 DB 181 GDPLIFGIALNLRTADIQMNSSNAQLLDYCSSKGYNISWELNEPNPSFKKADIFINS 240  
 QY 241 QLGSDYIQLHLKLRKSTFKNAKLYGPDVGQPRRTAKTAMLSFLKAGGEVIDSTWHYYL 300  
 DB 241 QLGSDYIQLHLKLRKSTFKNAKLYGPDVGQPRRTAKTAMLSFLKAGGEVIDSTWHYYL 300

QY 301 NGRTATREDPLNPVLDIFISSVOKFQVVESTBPKKVLGETSASAYGAGPALLSDTFA 360  
 DB 301 NGRTATREDPLNPVLDIFISSVOKFQVVESTBPKKVLGETSASAYGAGPALLSDTFA 360  
 QY 361 AGFWMLDLKLSAARWIEVNRQVFGAGNYHLVDENDDPLDYWLSILFFKKLVGTXYLM 420  
 DB 361 AGFWMLDLKLSAARWIEVNRQVFGAGNYHLVDENDDPLDYWLSILFFKKLVGTXYLM 420  
 QY 421 ASVQSKRRKRLRYVILHCTNTNPNRYKGGDLTLVAINLHNTKYRLRPFSNKQVDKYL 480  
 DB 421 ASVQSKRRKRLRYVILHCTNTNPNRYKGGDLTLVAINLHNTKYRLRPFSNKQVDKYL 480  
 QY 481 RPLGPHGLSSKSVQVNLGTLKQVDDQTLPLMEKPLRPGSSGLPAPFSYFFVIRNAKYA 540  
 DB 481 RPLGPHGLSSKSVQVNLGTLKQVDDQTLPLMEKPLRPGSSGLPAPFSYFFVIRNAKYA 540  
 QY 541 ACT 543  
 DB 541 ACT 543  
 RESULT 5  
 AA97635  
 ID AA97635 standard; Protein; 543 AA.  
 XX  
 AC AA97635;  
 XX  
 DT 20-APR-2001 (first entry)  
 XX  
 DE Human heparanase protein sequence.  
 XX  
 KW Heparanase; hnpbl; wound healing; angiogenesis; restenosis; Scurvy;  
 KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
 KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
 KW gene therapy; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200100643-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 19-JUN-2000; 2000MO-IL00358.  
 XX  
 PR 25-JUN-1999; 99US-0140801.  
 XX  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA  
 PA Pecker I, Michal I, Itzhaki H;  
 PI  
 XX WPI: 2001-137930/14.  
 DR  
 XX  
 PT New polynucleotides and polypeptides that are distantly homologous to  
 PT heparanase, useful in wound healing, as well as in gene therapy  
 PT protocols for angiogenesis, restenosis, atherosclerosis, or  
 PT inflammation -  
 XX  
 PS Disclosure; Page 64-65; 67pp; English.  
 XX  
 CC This sequence represents a heparanase of the invention.  
 CC The heparanase DNA and protein sequences are useful in wound healing,  
 CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary  
 CC disease, neurodegenerative diseases (such as Scurvy, Alzheimer's  
 CC disease, and Creutzfeldt-Jakob disease) or viral infections. The  
 CC heparanase coding sequence is particularly useful in gene therapy.  
 CC  
 SO Sequence 543 AA;  
 Query Match 100.0%; Score 2842; DB 22; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-274;  
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MLRSKRALPPMLLLGLPLSPGALPRPAQADVLDLFTQEPHLVSPSLSVT 60

DB 1 MLRSKRALPPMLLLGPISSRAIRPQAQDVDDDFTOEPLHLVSPSLST 60  
 QY 61 IDANLATDPPEFLILGSPKRLTLARGSPAYLRFGGKTDPFLIPDKKSTFEERSYMS 120  
 DB 61 IDANLATDPPEFLILGSPKRLTLARGSPAYLRFGGKTDPFLIPDKKSTFEERSYMS 120  
 QY 121 QVNODICKGSIIPDVEEKLRLMPYOBOLLRHYYOKKFNSTYSRSSVDVLYTFANCS 180  
 DB 121 QVNODICKGSIIPDVEEKLRLMPYOBOLLRHYYOKKFNSTYSRSSVDVLYTFANCS 180  
 QY 181 GDLIFGIALNLRTADLQWSSNAQLLDYCSKGYNISWEIGNEPNSFLKKADIFINGS 240  
 DB 181 GDLIFGIALNLRTADLQWSSNAQLLDYCSKGYNISWEIGNEPNSFLKKADIFINGS 240  
 QY 241 QLGEDYIQLHKLRLKSTFKNAKLYGPDVGQPRKRTAKMLKSLKAGGEVIDSVTHHHYLL 300  
 DB 241 QLGEDYIQLHKLRLKSTFKNAKLYGPDVGQPRKRTAKMLKSLKAGGEVIDSVTHHHYLL 300  
 QY 301 NGRTATREDPFLNDVDLFISSVQKVFQVVESTRPGKKVWLGETSAYGGAPLSDTPA 360  
 DB 301 NGRTATREDPFLNDVDLFISSVQKVFQVVESTRPGKKVWLGETSAYGGAPLSDTPA 360  
 QY 361 AGFMWLDKLGISARMGIEVVMROVFGAGNYHLVDENFDPDLYWLSLFLKLVGKVKLM 420  
 DB 361 AGFMWLDKLGISARMGIEVVMROVFGAGNYHLVDENFDPDLYWLSLFLKLVGKVKLM 420  
 QY 421 ASVQSKRRRLRYVLAHCTNTDNPYKEGDLTYAINLHNTKYLRPLPFPSNKQVDKYL 480  
 DB 421 ASVQSKRRRLRYVLAHCTNTDNPYKEGDLTYAINLHNTKYLRPLPFPSNKQVDKYL 480  
 QY 481 RPLGPHGLSKSVQVQNGLTLMVDDQTLPLMEKPLRPSSLGLPAFSYSFVIRNAKVA 540  
 DB 481 RPLGPHGLSKSVQVQNGLTLMVDDQTLPLMEKPLRPSSLGLPAFSYSFVIRNAKVA 540  
 QY 541 ACT 543  
 DB 541 ACT 543

RESULT 6  
 ABB07813  
 ID ABB07813 standard; protein; 543 AA.

AC ABB07813;  
 XX  
 DT 03-JUL-2002 (first entry)  
 XX  
 DS Human heparanase sequence.  
 XX  
 KM Heparanase; catalytic; cytosolic; antiviral; antibacterial; enzyme;  
 XX anti-protozoan; neuroprotective; heparin; human.  
 OS Homo sapiens.  
 XX  
 XX Key location/Qualifiers  
 XX Peptide 1..33  
 XX Protein /note= "signal peptide"  
 XX /note= "mature protein"  
 XX  
 XX US2002034810-A1.  
 XX  
 XX 21-MAR-2002.  
 XX  
 XX 16-AUG-2001; 2001US-0930218.  
 XX  
 XX 20-SEP-2000; 2000US-0666390.  
 XX  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 XX  
 XX Goldsmith O, Pecker I, Vlodevsky I, Michal I, Zcharia E;  
 XX

DR WPI: 2002-338926/37  
 XX Nucleic acid encoding avian and reptile heparanase polypeptide is  
 PT useful to treat various heparin-related disorders and the signal  
 PT peptide is useful in production of membrane-targeted or secreted  
 PT recombinant proteins  
 XX  
 XX Disclosure: Fig 1a; 3pp; English.  
 XX  
 XX The invention relates to an isolated avian and reptile nucleic acid,  
 CC encoding a polypeptide with heparanase catalytic activity. The signal  
 CC peptide of the nucleic acid can be used to express membrane-associated or  
 CC secreted proteins in heterologous expression systems. The encoded  
 CC polypeptides can be used to prevent tumor angiogenesis, metastasis and  
 CC invasion, and to intervene with pathologies associated with impaired  
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
 CC cellular susceptibility to viral, protozoa and bacterial infections or  
 CC disintegration of neurodegenerative plaques. The present sequence  
 CC represents a human heparanase protein sequence used in similarity  
 CC studies.  
 CC  
 CC Sequence 543 AA;  
 SQ  
 Query Match 100.0%; Score 2842; DB 23; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-274;  
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKRALPPMLLLGPISSRAIRPQAQDVDDDFTOEPLHLVSPSLST 60  
 DB 1 MLRSKRALPPMLLLGPISSRAIRPQAQDVDDDFTOEPLHLVSPSLST 60  
 QY 61 IDANLATDPPEFLILGSPKRLTLARGSPAYLRFGGKTDPFLIPDKKSTFEERSYMS 120  
 DB 61 IDANLATDPPEFLILGSPKRLTLARGSPAYLRFGGKTDPFLIPDKKSTFEERSYMS 120  
 QY 121 QVNODICKGSIIPDVEEKLRLMPYOBOLLRHYYOKKFNSTYSRSSVDVLYTFANCS 180  
 DB 121 QVNODICKGSIIPDVEEKLRLMPYOBOLLRHYYOKKFNSTYSRSSVDVLYTFANCS 180  
 QY 181 GDLIFGIALNLRTADLQWSSNAQLLDYCSKGYNISWEIGNEPNSFLKKADIFINGS 240  
 DB 181 GDLIFGIALNLRTADLQWSSNAQLLDYCSKGYNISWEIGNEPNSFLKKADIFINGS 240  
 QY 241 QLGEDYIQLHKLRLKSTFKNAKLYGPDVGQPRKRTAKMLKSLKAGGEVIDSVTHHHYLL 300  
 DB 241 QLGEDYIQLHKLRLKSTFKNAKLYGPDVGQPRKRTAKMLKSLKAGGEVIDSVTHHHYLL 300  
 QY 301 NGRTATREDPFLNDVDLFISSVQKVFQVVESTRPGKKVWLGETSAYGGAPLSDTPA 360  
 DB 301 NGRTATREDPFLNDVDLFISSVQKVFQVVESTRPGKKVWLGETSAYGGAPLSDTPA 360  
 QY 361 AGFMWLDKLGISARMGIEVVMROVFGAGNYHLVDENFDPDLYWLSLFLKLVGKVKLM 420  
 DB 361 AGFMWLDKLGISARMGIEVVMROVFGAGNYHLVDENFDPDLYWLSLFLKLVGKVKLM 420  
 QY 421 ASVQSKRRRLRYVLAHCTNTDNPYKEGDLTYAINLHNTKYLRPLPFPSNKQVDKYL 480  
 DB 421 ASVQSKRRRLRYVLAHCTNTDNPYKEGDLTYAINLHNTKYLRPLPFPSNKQVDKYL 480  
 QY 481 RPLGPHGLSKSVQVQNGLTLMVDDQTLPLMEKPLRPSSLGLPAFSYSFVIRNAKVA 540  
 DB 481 RPLGPHGLSKSVQVQNGLTLMVDDQTLPLMEKPLRPSSLGLPAFSYSFVIRNAKVA 540  
 QY 541 ACT 543  
 DB 541 ACT 543

RESULT 7  
 AAY02346  
 ID AAY02346 standard; Protein; 592 AA.

XX

AC AAY02346;  
 XX  
 DT 09-JUL-1999 (first entry)  
 XX  
 DE A human heparanase protein.  
 XX  
 KW Heparanase; hpa; modulator; heparin-binding growth factor;  
 KW cellular response; cytokine; cell interaction; plasma lipoprotein;  
 KW cellular susceptibility; infection; disintegration;  
 KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;  
 KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;  
 KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.  
 XX  
 OS Homo-sapiens.  
 XX  
 EN MO9111798-A1.  
 XX  
 PD 11-MAR-1999.  
 XX  
 PF 31-AUG-1998; 98MO-US17954.  
 XX  
 PR 02-JUL-1998; 98US-0109386.  
 PR 02-SEP-1997; 97US-0922170.  
 XX  
 XX (FRIE/) FRIEDMAN M M.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA (INST-) INSIGHT STRATEGY & MARKETING LTD.  
 XX  
 PI Feinstein E, Pecker I, Vlodavsky I;  
 XX  
 DR WPI: 1999-302255/25.  
 DR N-PSDB; AAA35650.  
 XX  
 PT New human polynucleotide useful for treating angiogenesis,  
 PT restenosis, and inflammation  
 XX  
 XX Claim 6; Page 65-66; 63pp; English.  
 XX  
 CC The specification describes a polypeptide having heparanase (hpa)  
 CC activity. The recombinant protein is used as a modulator of  
 CC heparin-binding growth factors, cellular responses to heparin-binding  
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
 CC cellular susceptibility to viral, protozoal and bacterial infections  
 CC or disintegration of neurodegenerative plaques. Heparanase may be  
 CC useful for conditions such as wound healing, angiogenesis, restenosis,  
 CC atherosclerosis, inflammation, neurodegenerative diseases, and viral  
 CC infections. Mammalian heparanase can be used to neutralize plasma  
 CC heparin, and anti-heparanase antibodies may be applied for  
 CC immunodetection and diagnosis of micrometastases, autoimmune lesions,  
 CC and renal failure in biopsy specimens, plasma samples, and body fluids.  
 CC The present sequence represents human heparanase.  
 XX  
 XX Sequence 592 AA;  
 XX  
 Query Match 100.0%; Score 2842; DB 20; Length 592;  
 Best Local Similarity 100.0%; Pred. No. 1,1e-273; Indels 0; Gaps 0;  
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MLRSKALPPPLMLLGLPGLSPALPRPAQCVVDLDFPTQEPHLVSPSFLSVT 60  
 DB 50 MLRSKALPPPLMLLGLPGLSPALPRPAQCVVDLDFPTQEPHLVSPSFLSVT 109  
 QY 61 IDANLATDPPFLILGSPKRTTLARGISPAVYLRFGTKTDFLFDPKCESTFEBSRYWOS 120  
 DB 110 IDANLATDPPFLILGSPKRTTLARGISPAVYLRFGTKTDFLFDPKCESTFEBSRYWOS 169  
 QY 121 QNVADICKYSIPDVEXKRLRLEMPYOBOLLRHYOKKFKNSYSSSVVLYTFPANC 180  
 DB 170 QNVADICKYSIPDVEXKRLRLEMPYOBOLLRHYOKKFKNSYSSSVVLYTFPANC 229  
 QY 181 GLDLIFGLINLRLTRADLQNNSSNAQLLDVCCSGYNIISWELGNEPNSFLKADIFPINGS 240  
 DB 230 GLDLIFGLINLRLTRADLQNNSSNAQLLDVCCSGYNIISWELGNEPNSFLKADIFPINGS 289

QY 241 QUGEDYIOLHLKLRKSTFKNAKLYGPDVGOBRRTAKMLKSLKAGCEVIDSVTHHYLL 300  
 DB 290 QUGEDYIOLHLKLRKSTFKNAKLYGPDVGOBRRTAKMLKSLKAGCEVIDSVTHHYLL 349  
 QY 301 NGRTATREDPLNPVDLIDFISSVQKVEQVVESTPRGKVMIGETSSAYGGAPELLSDTFA 360  
 DB 350 NGRTATREDPLNPVDLIDFISSVQKVEQVVESTPRGKVMIGETSSAYGGAPELLSDTFA 409  
 QY 361 AGFWMLDKLGLSARNGIEVVRQVFPAGNYHLVDENFDPLDYWLSLLEFKLVGTQKVM 420  
 DB 410 AGFWMLDKLGLSARNGIEVVRQVFPAGNYHLVDENFDPLDYWLSLLEFKLVGTQKVM 469  
 QY 421 ASVQSKRKRLRVYHACNTNTPRYKEGDLTLVYALINLNTKLYRLPYPFANKOVDKYLL 480  
 DB 470 ASVQSKRKRLRVYHACNTNTPRYKEGDLTLVYALINLNTKLYRLPYPFANKOVDKYLL 529  
 QY 481 RPLGPHGLSKSVQVQNGTLTKMVDQTLPLMEKPLRPGSSILGLPAPSYSPFVIRNAKVA 540  
 DB 530 RPLGPHGLSKSVQVQNGTLTKMVDQTLPLMEKPLRPGSSILGLPAPSYSPFVIRNAKVA 589  
 QY 541 ACI 543  
 DB 590 ACI 592  
 RESULT 8  
 AAB08850  
 ID AAB08850 standard; Protein; 592 AA.  
 XX  
 XX AAB08850;  
 AC  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX  
 XX Amino acid sequence of a human heparanase polypeptide.  
 DE  
 XX  
 KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;  
 KW heparin-binding growth factor; cytokine; neurodegenerative plaque;  
 KW wound healing; infection; burn; angiogenesis; restenosis;  
 KW atherosclerosis; inflammation; neurodegenerative disease;  
 KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 PN MO20005178-A1.  
 PD 08-SEP-2000.  
 XX  
 PF 14-FEB-2000; 2000MO-US03542.  
 XX  
 PR 01-MAR-1999; 99US-0258892.  
 XX  
 XX (INST-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV  
 PA (FRIE/) FRIEDMAN M M.  
 PI Pecker I, Vlodavsky I, Feinstein E;  
 XX  
 XX WPI: 2000-579289/54.  
 DR N-PSDB; AAA75053.  
 XX  
 PT New polynucleotides encoding a polypeptide having heparanase activity,  
 PT useful in wound healing and in gene therapy, particularly in treating  
 PT tumour, inflammation, autoimmunity, neurodegenerative diseases  
 XX  
 XX Claim 22; Page 122-123; 152pp; English.  
 XX  
 CC The present sequence represents a human protein with heparanase catalytic  
 CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,  
 CC particularly in treating tumour, inflammation or autoimmunity.  
 CC Particularly, the polynucleotide is useful in modulating the  
 CC bioavailability of heparin-binding growth factors, cellular responses  
 CC to heparin-binding growth factors (e.g. bFGF) and cytokines

(e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular susceptibility to certain viral and some bacterial and protozoa infections, or disintegration of neurodegenerative plaques. The polynucleotide is also useful in wound healing (e.g. thermal, chemical or radiation burns), and in the treatment of angiogenesis, restenosis, atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral, bacterial or protozoa infections.

Sequence 592 AA;

Query Match 100.0%; Score 2842; DB 21; Length 592;

Best Local Similarity 100.0%; Pred. No. 1.1e-273;

Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 MLRSKRALPPMLLLGLPLGSPGALPRPAQADVDVDFPQEPHLVSPFLSVT 60  
50 MLRSKRALPPMLLLGLPLGSPGALPRPAQADVDVDFPQEPHLVSPFLSVT 109  
DB 61 IDANLATDPPRLILGSPKRLTARGLSPAYLRFPGTKTDFLFDPKKSTFEERSYMO 120  
110 IDANLATDPPRLILGSPKRLTARGLSPAYLRFPGTKTDFLFDPKKSTFEERSYMO 169  
DB 121 QVNODICKKGSIPDVBEKRLRLEMPYQEOQLLREHYOKKFNSTYSRSSVDLYTFANCS 180  
170 QVNODICKKGSIPDVBEKRLRLEMPYQEOQLLREHYOKKFNSTYSRSSVDLYTFANCS 229  
DB 181 GDLLIFGALNALRTADLQWSSNAQLLDYCSSKGYNISWELGNENPNSFLKADIFINGS 240  
230 GDLLIFGALNALRTADLQWSSNAQLLDYCSSKGYNISWELGNENPNSFLKADIFINGS 289  
DB 241 QUGEDYIQLHLKLRKSTFNKAKLYGPDVQGPRRKTAAMLKSLFKAGGEVIDSVTHHHYLL 300  
290 QUGEDYIQLHLKLRKSTFNKAKLYGPDVQGPRRKTAAMLKSLFKAGGEVIDSVTHHHYLL 349  
DB 301 NGRTATREDPLNPVDLIFISSVQKVFQVVESTPRGKVMKLGSTSSAYGGAPLLSDTFA 360  
350 NGRTATREDPLNPVDLIFISSVQKVFQVVESTPRGKVMKLGSTSSAYGGAPLLSDTFA 409  
DB 361 AGFPMWLKGLSARMGIEVVMROVFFGAGNYHLVDENFDPLPDYMLSLFLFKLVGTKVL 420  
410 AGFPMWLKGLSARMGIEVVMROVFFGAGNYHLVDENFDPLPDYMLSLFLFKLVGTKVL 469  
DB 421 ASVQSKRRKRLRVYLHCTNTDNRPRYKGGDLTYAINHNTKYLRLPYPSNKOVDKYL 480  
470 ASVQSKRRKRLRVYLHCTNTDNRPRYKGGDLTYAINHNTKYLRLPYPSNKOVDKYL 529  
DB 481 RPLGPPGLLSKSYQVNLGLTKMVDOTLPLMEKPLRPGSSLGLPAFSYSFPVIRNAKVA 540  
530 RPLGPPGLLSKSYQVNLGLTKMVDOTLPLMEKPLRPGSSLGLPAFSYSFPVIRNAKVA 589  
DB 541 ACT 543  
590 ACT 592

RESULT 9

ID AAY17082 standard; Protein; 543 AA.

AC AAY17082;

DT 21-JUL-1999 (first entry)

DE Human heparanase enzyme.

KX Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;  
KW metacastais; angiogenesis; wound healing; angioplasty-induced restenosis;  
KW arteriosclerosis; atherosclerosis; inflammation; tissue development;  
human; HSPG.

OS Homo sapiens.

XX

PN MO9921975-21.  
XX 06-MAY-1999.  
PD 28-OCT-1998; 98MO-AU00898.  
PF 09-DEC-1997; 97AU-0000812.  
PR 28-OCT-1997; 97AU-0000062.  
XX (AUSU) UNIV AUSTRALIAN NAT.

PI Freeman CG, Hamdorf BJ, Hulett MD, Parish CR;  
XX WPI, 1999-312956/26.  
DR N-PSDB; MAX37259.

PT Polynucleotides encoding mammalian endoglucuronidases, especially  
XX heparanases, useful to promote wound healing  
XX Claim 6; Page 69-73; 112pp; English.

PS The invention relates to nucleic acid sequences that encode heparanase  
XX enzymes having endoglucuronidase activity. Recombinant heparanases are  
XX capable of removing the HS side chain from heparan sulfate proteoglycan  
XX (HSPG). Sulfated oligosaccharides, sulfonates or HSPG can be used to  
XX inhibit heparanase, this is useful for treatment of a physiological or  
XX medical condition associated with elevated heparanase activity, such as  
XX metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,  
XX arteriosclerosis, atherosclerosis and inflammation. The human, murine and  
XX rat heparanases can be used to enhance wound healing, especially  
XX associated with tissue development and repair. The conditions mentioned  
XX above can be diagnosed using specific antibodies, and also using primers  
XX and probes specific for the heparanase polynucleotides. Other uses of the  
XX heparanases include sequencing sulfated molecules such as HSPG. The  
XX present sequence represents a human heparanase.

Sequence 543 AA;

Query Match 99.9%; Score 2838; DB 20; Length 543;

Best Local Similarity 99.8%; Pred. No. 2.4e-273;

Matches 542; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 1 MLRSKRALPPMLLLGLPLGSPGALPRPAQADVDVDFPQEPHLVSPFLSVT 60  
1 MLRSKRALPPMLLLGLPLGSPGALPRPAQADVDVDFPQEPHLVSPFLSVT 60  
DB 61 IDANLATDPPRLILGSPKRLTARGLSPAYLRFPGTKTDFLFDPKKSTFEERSYMO 120  
120 IDANLATDPPRLILGSPKRLTARGLSPAYLRFPGTKTDFLFDPKKSTFEERSYMO 120  
DB 121 QVNODICKKGSIPDVBEKRLRLEMPYQEOQLLREHYOKKFNSTYSRSSVDLYTFANCS 180  
170 QVNODICKKGSIPDVBEKRLRLEMPYQEOQLLREHYOKKFNSTYSRSSVDLYTFANCS 229  
DB 181 GDLLIFGALNALRTADLQWSSNAQLLDYCSSKGYNISWELGNENPNSFLKADIFINGS 240  
181 GDLLIFGALNALRTADLQWSSNAQLLDYCSSKGYNISWELGNENPNSFLKADIFINGS 240  
DB 241 QUGEDYIQLHLKLRKSTFNKAKLYGPDVQGPRRKTAAMLKSLFKAGGEVIDSVTHHHYLL 300  
290 QUGEDYIQLHLKLRKSTFNKAKLYGPDVQGPRRKTAAMLKSLFKAGGEVIDSVTHHHYLL 349  
DB 301 NGRTATREDPLNPVDLIFISSVQKVFQVVESTPRGKVMKLGSTSSAYGGAPLLSDTFA 360  
350 NGRTATREDPLNPVDLIFISSVQKVFQVVESTPRGKVMKLGSTSSAYGGAPLLSDTFA 409  
DB 361 AGFPMWLKGLSARMGIEVVMROVFFGAGNYHLVDENFDPLPDYMLSLFLFKLVGTKVL 420  
410 AGFPMWLKGLSARMGIEVVMROVFFGAGNYHLVDENFDPLPDYMLSLFLFKLVGTKVL 469  
DB 421 ASVQSKRRKRLRVYLHCTNTDNRPRYKGGDLTYAINHNTKYLRLPYPSNKOVDKYL 480  
470 ASVQSKRRKRLRVYLHCTNTDNRPRYKGGDLTYAINHNTKYLRLPYPSNKOVDKYL 529



QY 481 RPLGPHGLSKSVQVNGLTLMKVDQDTLPPLMEKPLRPSSGLGLPAFSSFFVIRNAKYA 540  
DB 481 RPLGPHGLSKSVQVNGLTLMKVDQDTLPPLMEKPLRPSSGLGLPAFSSFFVIRNAKYA 540  
QY 541 ACT 543  
DB 541 ACT 543

RESULT 10  
AAB86206 standard; Protein; 543 AA.  
AAB86206:  
24-AUG-2001 (first entry)  
Human heparanase inhibitor protein.  
Heparanase; inhibitor; cardiac insufficiency; cardiast; nephrotropic;  
hepatocytic; veterinary medicine; congestive heart failure; dyspnoea;  
primary cardiomyopathy; peripheral edema; pulmonary congestion;  
hepatic congestion; hydrothorax; ascites; nocturia; human.  
Homo sapiens.  
DE19955803-A1.  
23-MAY-2001.  
19-NOV-1999; 99DE-1055803.  
19-NOV-1999; 99DE-1055803.  
(KNOL) KNOLL AG.  
Herr D, Hahn A, Laux V;  
WPI; 2001-368371/39.  
N-PSDB; AAH20940.  
Treatment or prevention of cardiac insufficiency and dyspnoea, comprises  
administration of heparanase inhibitor  
disclosure; Page 11-13; 16pp; German.  
This invention describes a novel heparanase inhibitor which can be used  
for the treatment or prevention of cardiac insufficiency and associated  
indications, symptoms and/or malfunctions. The heparanase inhibitor of  
the invention has cardiast, nephrotropic and hepatocytic activity. The  
products of the invention can be used in human and veterinary medicine,  
for the treatment or prevention of congestive heart failure e.g. primary  
cardiomyopathy. Associated conditions treated or prevented with the  
inhibitor are especially peripheral edemas, pulmonary and hepatic  
congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.  
nocturia can also be treated. This sequence represents the human  
heparanase protein described in the method of the invention.

Sequence 543 AA;  
Query Match 99.94; Score 2838; DB:22; Length 543;  
Best Local Similarity 99.84; Pred. No. 2.4e-273;  
Matches 542; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKPLAPPPMLLLGLGLPLSPALPRPAQADVDVLDFTQEPHLVSPFLSYT 60  
DB 1 MLRSKPLAPPPMLLLGLGLPLSPALPRPAQADVDVLDFTQEPHLVSPFLSYT 60  
QY 61 IDANLATDPRFLILGSPKRTTLARGSPAYLRFPGTGTDFLLIDPKKSTFEERSYWS 120  
DB 61 IDANLATDPRFLILGSPKRTTLARGSPAYLRFPGTGTDFLLIDPKKSTFEERSYWS 120

QY 121 QVNOGICCKGSIPEVEEKLRLMEPYQQLREHYCKFNSTSRSSVDVLTFFANC 180  
DB 121 QVNOGICCKGSIPEVEEKLRLMEPYQQLREHYCKFNSTSRSSVDVLTFFANC 180  
QY 181 GDLIFGALNALLRTADLQWSSNAQLLDYSSKGYNISWELGNEPNSFLKKADIFING 240  
DB 181 GDLIFGALNALLRTADLQWSSNAQLLDYSSKGYNISWELGNEPNSFLKKADIFING 240  
QY 241 QGDEYIQLKLLKSTFRNAKLVPDVGGPRRTAKLKSFLRAGGVISVTMHHYLL 300  
DB 241 QGDEYIQLKLLKSTFRNAKLVPDVGGPRRTAKLKSFLRAGGVISVTMHHYLL 300  
QY 301 NGRTATREDPLNPVDLFISSVQKVPQVSTRGKVMGESSAYAGGAPLLSDTFA 360  
DB 301 NGRTATREDPLNPVDLFISSVQKVPQVSTRGKVMGESSAYAGGAPLLSDTFA 360  
QY 361 AGFWMLDKLGLSARNGIEVVRQVFFGAGNYHLDENFDPLDYWLSLLFKLVGTXYLM 420  
DB 361 AGFWMLDKLGLSARNGIEVVRQVFFGAGNYHLDENFDPLDYWLSLLFKLVGTXYLM 420  
QY 421 ASVQSKRKRLRVYLHCTNTDNPVKEDLTLXANLNNYKYLRLPPFSNKQVDKYL 480  
DB 421 ASVQSKRKRLRVYLHCTNTDNPVKEDLTLXANLNNYKYLRLPPFSNKQVDKYL 480  
QY 481 RPLGPHGLSKSVQVNGLTLMKVDQDTLPPLMEKPLRPSSGLGLPAFSSFFVIRNAKYA 540  
DB 481 RPLGPHGLSKSVQVNGLTLMKVDQDTLPPLMEKPLRPSSGLGLPAFSSFFVIRNAKYA 540  
QY 541 ACT 543  
DB 541 ACT 543

RESULT 11  
AAY30124 standard; Protein; 588 AA.  
AAY30124:  
20-MAR-2003 (updated)  
14-OCT-1999 (first entry)  
A human protein with heparanase activity.  
Human; heparanase; heparan sulfate; trauma; autoimmune disease;  
skin disease; cardiovascular disease; nervous system disease;  
Alzheimer's disease; cancer; cancer metastasis; angiogenesis;  
inflammation; arthritis.  
Homo sapiens.  
W09940207-A1.  
12-AUG-1999.  
05-FEB-1999; 95WO-BP00777.  
09-FEB-1998; 98GB-0002725.  
(NOVS) NOVARTIS AG.  
(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
Nakajima M, Toyoshima M;  
WPI; 1999-494300/41.  
N-PSDB; AAX86671.  
New heparanase polypeptide useful for treating autoimmune diseases,  
skin diseases, cardiovascular diseases and nervous system diseases  
including Alzheimer's disease  
Claim 3; Page 29-31; 40pp; English.

XX The present sequence represents a polypeptide with human heparanase  
CC biological activity. Antagonists and inhibitors of the protein prevent  
CC it from degrading the extracellular matrix and releasing heparan sulfate  
CC from the extracellular matrix surface. The heparanase protein or the  
CC anti-heparanase antibody are used in pharmaceutical compositions for  
CC treating warm blooded animals suffering from a disease resulting from  
CC shortage or lack of the heparanase protein, or from excessive activity  
CC or over-expression of the heparanase protein, respectively. The  
CC heparanase protein is used in treating diseases such as trauma,  
CC autoimmune disease, skin diseases, cardiovascular diseases and nervous  
CC system diseases including Alzheimer's disease resulting from shortage or  
CC lack of polypeptide. The anti-heparanase antibody is used in treating  
CC the diseases like cancer, cancer metastasis, angiogenesis and  
CC inflammation including arthritis resulting from excessive activity or  
CC over expression of heparanase protein. The anti-heparanase antibody can  
CC be used to detect the presence or absence of polypeptide and its  
CC concentration.  
CC (updated on 20-MAR-2003 to correct PA field.)

**SQ Sequence 588 AA;**

Query Match	99.9%	Score 2838	DB 20	Length 588
Best Local Similarity	99.8%	Pred. No. 2.7e-273		
Matches 542	Conservative	1	Mismatches 0	Indels 0

Qy	I	MULRSKDALPPLPMLMLLLGLGLPGLPSGALPPRAQAOVDVDDLDFOTOPRLHVSSFFLSVT	60
Db	46	MULRSKDALPPLPMLMLLLGLGLPGLPSGALPPRAQAOVDVDDLDFOTOBRLHVSSFFLSVT	105
Qy	61	IDANLATDPRFLILGSPPLRTLARGLSAPYLRFPGTITDELIDDPKKESTFEERSYWG	120
Db	106	IDANLATDPRFLILGSPPLRTLARGLSAPYLRFPGTITDELIDDPKKESTFEERSYWG	165
Qy	121	QVNOIDICTGSIPPDVEEKLRLLEMPYOBOLLIREHYOKKFPKNSYSSNSVDVLTTPANC	180
Db	166	QVNOIDICTGSIPPDVEEKLRLLEMPYOBOLLIREHYOKKFPKNSYSSNSVDVLTTPANC	225
Qy	181	GUDLIRFGNALARTADLOMNSSNAOLLDDYCSKGYNI SWELEGNBPNSFLKKADI PINGS	240
Db	226	GUDLIRFGNALARTADLOMNSSNAOLLDDYCSKGYNI SWELEGNBPNSFLKKADI PINGS	285
Qy	241	QLGDEDYIOHLKLRKSTFPNAKLVPBDVQOPPRKAKAKLKSFLRAGSEVIDSVTHHHYLL	300
Db	286	QLGDEDYIOHLKLRKSTFPNAKLXPBDVQOPPRKAKAKLSFLRAGSEVIDSVTHHHYLL	345
Qy	301	NGRTATREDPLNPDVLDIFISSVQKVPQVVESTRPCKYWLGETSSAYGCGAPLLSDTFA	360
Db	346	NGRTATREDPLNPDVLDIFISSVQKVPQVVESTRPCKYWLGETSSAYGCGAPLLSDTFA	405
Qy	361	AGTMMWLDKGLSARMKI EYVMRQVFFGAGNHLVDENBDPLPDYWL SLLPFKL VGTNYLM	420
Db	406	AGTMMWLDKGLSARMKI EYVMRQVFFGAGNHLVDENBDPLPDYWL SLLPFKL VGTNYLM	465
Qy	421	ASVQSGSKRRKLRYVYLHCTNDNPRKXEGDLTYAINTLANTKYRLLPYPSNKOVDKCYLL	480
Db	466	ASVQSGSKRRKLRYVYLHCTNDNPRKXEGDLTYAINTLANTKYRLLPYPSNKOVDKCYLL	525
Qy	481	RPLGPHGLLSKSYVOLNGLTLKXVDDQUTLEPLMEKELRPGSSLSGLPASYSPFVIRNAKVA	540
Db	526	RPLGPHGLLSKSYVOLNGLTLKXVDDQUTLEPLMEKELRPGSSLSGLPASYSPFVIRNAKVA	585
Qy	541	ACI 543	
Db	586	ACI 588	

DT 23-MAY-2001 (first entry)

DE Human membrane or secretory protein clone P58C0090.  
XX  
XX  
XX Human; secretory protein; membrane protein; vaccine; gene therapy.  
KW rheumatoid arthritis; diabetes

OS Homo sapiens.

PN EP1067182-A2

PD 10-JAN-2001.

PF 07-JUL-2000; 2000EP-0114090.

PR 08-JUL-1999; 99JP-0194179.

PR 02-MAY-2000; 2000JP-0183766.

PA (HELI-) HELIX RES INST.  
xx

PI Ota T, Isogai T, Nishikawa

DR WPI; 2001-093989/11.

XX  
XX

PT gene therapy or as candidate

PS Claim 1; SEQ ID 90; 609bp +  
yy

CC This invention relates to nu  
CC which exceeds human secretory

CC AAB88311 AAB88419. Included in the invention are primers  
CC AAF63917 AAF64295 and AAF62232 - AAF62235 which are used to isolate the  
CC cDNA sequences of the invention. The invention also includes methods for  
CC the production of antibodies directed against the proteins, and cDNA  
CC sequences, which can be used in vaccines. The polynucleotide sequences  
CC can be used in gene therapy. The polynucleotide sequences and the  
CC proteins they encode may be used in the prevention, treatment and  
CC diagnosis of diseases associated with inappropriate secretory  
CC protein/membrane protein expression. The nucleic acids and complementary  
CC sequences may also be used as DNA probes in diagnostic assays  
CC (e.g. polymerase chain reactions (PCR)) to detect and quantify the  
CC presence of similar nucleic acid sequences in samples. They may also be  
CC used to study the expression and function of secretory proteins/membrane  
CC polypeptides and their role in metabolism. The polypeptides may be used  
CC as antigens in the production of antibodies against them and in assays to  
CC identify modulators (agonists and antagonists) of expression and  
CC activity. The antibodies and antagonists may also be used as therapeutic.  
CC agents to down regulate expression and activity. The antibodies may also  
CC be used as diagnostic agents for detecting the presence of the  
CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay  
CC (ELISA)). Examples of diseases which may be treated include rheumatoid  
CC arthritis and diabetes.

**SQ**      **Sequence**      **543 AA;**

Query Match	99.4%	Pred. 2826	DB 22	Length 543
Best Local Similarity	99.4%	Pred. No. 3.8e-272		
Matches 540	Conservative	2	Mismatches 1	Indels 0
				Gaps 0

QY	1	MLIRSGPALPPPLMLLLGLPGPLSPGALLPPRPAQADVDLDFOTQRPPLHVSSFLSVT	60
Db	1	MLIRSGPALPPPLMLLLGLPGPLSPGALLPPRPAQADVDLDFOTQRPPLHVSSFLSVT	60
QY	61	IDANLATDPRFILLGSPKRLTARGSPAYLRGCKTDFLIDPKKESFPERBSTWQS	120
Db	61	IDANLATDPRFILLGSPKRLTARGSPAYLRGCKTDFLIDPKKESFPERBSTWQS	120
QY	121	QVNVQDLCCKGSLPQVVEEKLRLIEWYQQLLRHYOKKFNSTYSRSSVDVLTYPNCS	180
Db	121	QVNVQDLCCKGSLPQVVEEKLRLIEWYQQLLRHYOKKFNSTYSRSSVDVLTYPNCS	180

QY 181 GDLITGILNALRTADLQNNSSNAQLLDYCSKGYNISMEIGNEPSFLKADIFIN 240  
 DB 181 GDLITGILNALRTADLQNNSSNAQLLDYCSKGYNISMEIGNEPSFLKADIFIN 240  
 QY 241 QLGEDYIQLHLKLRKSTFKNAKLYGPDVGPARRKTAAMLKSPFKAGGEVIDSVTHMYL 300  
 DB 241 QLGEDYIQLHLKLRKSTFKNAKLYGPDVGPARRKTAAMLKSPFKAGGEVIDSVTHMYL 300  
 QY 301 NGRTATREDPLNDVDLIFISSVQKVFQVESTRPCKKWLGETSSAYGGAPLSDTFA 360  
 DB 301 NGRTATREDPLNDVDLIFISSVQKVFQVESTRPCKKWLGETSSAYGGAPLSDTFA 360  
 QY 361 AGPMWLDKLGSLRMGIEVVMROVFGAGNYHLVDENPDLDPYWLSTLFLKLVGKTKV 420  
 DB 361 AGPMWLDKLGSLRMGIEVVMROVFGAGNYHLVDENPDLDPYWLSTLFLKLVGKTKV 420  
 QY 421 ASVQSKRRKRLRYLHCTNTDNPARYEGDLTLVAINLHVNTKYLRLPYFSNKOVDKYL 480  
 DB 421 ASVQSKRRKRLRYLHCTNTDNPARYEGDLTLVAINLHVNTKYLRLPYFSNKOVDKYL 480  
 QY 481 RPLGPHGLLSKSVQNLGLTLKAVDDOTLPLMEKPLRPSSISGLPFAFSYFFVIRNAKYA 540  
 DB 481 RPLGPHGLLSKSVQNLGLTLKAVDDOTLPLMEKPLRPSSISGLPFAFSYFFVIRNAKYA 540  
 QY 541 ACT 543  
 DB 541 ACT 543

## RESULT 13

ABPS6822  
 ID ABPS6822 standard; Protein; 545 AA.

AC ABPS6822;  
 DT 02-APR-2003 (first entry)  
 XX  
 DE Human heparanase protein SEQ ID NO:18.  
 XX  
 KW Human; heparanase; phosphotriphosphate; antilysine oligonucleotide;  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003004705-A1.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PP 01-JUL-2002; 2002WO-US20636.  
 XX  
 PR 05-JUL-2001; 2001US-0899440.  
 XX  
 PA (UYCO) UNIV COLUMBIA NEW YORK.  
 XX  
 PI Stein C;  
 XX  
 DR WPI; 2003-201558/19.  
 XX  
 DR N-PSDB; AB222816.  
 XX  
 PT New oligonucleotide having a sequence complementary to a sequence of  
 PT ribonucleic acid encoding a heparanase, useful for preparing a  
 PT composition for treating tumor.  
 XX  
 PS Disclosure; Page 46-47; 48pp; English.

CC The present invention describes an oligonucleotide having a sequence  
 CC complementary to a sequence of ribonucleic acid encoding a heparanase.  
 CC The oligonucleotide hybridizes with the ribonucleic acid under conditions  
 CC of high stringency and has a sequence comprising 10-40 bp. The  
 CC internucleotide linkages of the oligonucleotide comprise at least one  
 CC phosphorothioate linkage. Hybridisation of the oligonucleotide to the  
 CC ribonucleic acid inhibits expression of the heparanase, where inhibition

CC of heparanase means at least a 50% reduction in the quality of  
 CC heparanase. Also described: (1) a method of inhibiting expression of a  
 CC heparanase in a cell; (2) a composition comprising the above  
 CC oligonucleotide in an amount effective to inhibit the expression of  
 CC heparanase in the cell and a carrier; and (3) a method of treating a  
 CC tumor in a subject comprising administering to the subject an amount of  
 CC the above oligonucleotide effective to inhibit expression of a heparanase  
 CC in the subject. Heparanase antisense oligonucleotides have cytostatic  
 CC activity, can be used in gene therapy, and can be used for preparing a  
 CC composition for treating tumors. The present sequence represents human  
 CC heparanase, which is given in the exemplification of the present  
 CC invention.

CC Sequence 545 AA;

Query Match 99.1%; Score 2817; DB 24; Length 545;  
 Best Local Similarity 99.4%; Pred. No. 3e-771;  
 Matches 542; Conservative 1; Mismatches 0; Indels 2; Gaps 2;

QY 1 MLNLSKRALPP-IMLILGGLSGALPRPAQA-ODVVDLDFPOEPLHLVSPFLS 58  
 DB 1 MLNLSKRALPP-IMLILGGLSGALPRPAQA-ODVVDLDFPOEPLHLVSPFLS 60  
 QY 59 VTIDANLATDRPFLILGSPKLTARGLSPAYLRFSGTKDFLIPDKKSTEEERSY 118  
 DB 59 VTIDANLATDRPFLILGSPKLTARGLSPAYLRFSGTKDFLIPDKKSTEEERSY 120  
 QY 61 VTIDANLATDRPFLILGSPKLTARGLSPAYLRFSGTKDFLIPDKKSTEEERSY 120  
 DB 61 VTIDANLATDRPFLILGSPKLTARGLSPAYLRFSGTKDFLIPDKKSTEEERSY 120  
 QY 119 OSQVNOICIKYGSIPDVEEKLRLWPYQEQQLRLREHYOKKFNKSTYSRSSVDVLYTFAN 178  
 DB 119 OSQVNOICIKYGSIPDVEEKLRLWPYQEQQLRLREHYOKKFNKSTYSRSSVDVLYTFAN 180  
 QY 121 OSQVNOICIKYGSIPDVEEKLRLWPYQEQQLRLREHYOKKFNKSTYSRSSVDVLYTFAN 180  
 DB 121 OSQVNOICIKYGSIPDVEEKLRLWPYQEQQLRLREHYOKKFNKSTYSRSSVDVLYTFAN 180  
 QY 179 CSGLDITGILNALRTADLQNNSSNAQLLDYCSKGYNISMEIGNEPSFLKADIFIN 238  
 DB 179 CSGLDITGILNALRTADLQNNSSNAQLLDYCSKGYNISMEIGNEPSFLKADIFIN 240  
 QY 239 GSQLGEDYIQLHLKLRKSTFKNAKLYGPDVGPARRKTAAMLKSPFKAGGEVIDSVTHMY 298  
 DB 239 GSQLGEDYIQLHLKLRKSTFKNAKLYGPDVGPARRKTAAMLKSPFKAGGEVIDSVTHMY 300  
 QY 241 GSQLGEDYIQLHLKLRKSTFKNAKLYGPDVGPARRKTAAMLKSPFKAGGEVIDSVTHMY 300  
 DB 241 GSQLGEDYIQLHLKLRKSTFKNAKLYGPDVGPARRKTAAMLKSPFKAGGEVIDSVTHMY 300  
 QY 299 YLNGRTATREDPLNDVDLIFISSVQKVFQVESTRPCKKWLGETSSAYGGAPLSDT 358  
 DB 299 YLNGRTATREDPLNDVDLIFISSVQKVFQVESTRPCKKWLGETSSAYGGAPLSDT 360  
 QY 301 YLNGRTATREDPLNDVDLIFISSVQKVFQVESTRPCKKWLGETSSAYGGAPLSDT 360  
 DB 301 YLNGRTATREDPLNDVDLIFISSVQKVFQVESTRPCKKWLGETSSAYGGAPLSDT 360  
 QY 359 FAAGFMWLDKLGSLRMGIEVVMROVFGAGNYHLVDENPDLDPYWLSTLFLKLVGKTKV 418  
 DB 359 FAAGFMWLDKLGSLRMGIEVVMROVFGAGNYHLVDENPDLDPYWLSTLFLKLVGKTKV 420  
 QY 419 LMASVQSKRRKRLRYLHCTNTDNPARYEGDLTLVAINLHVNTKYLRLPYFSNKOVDKYL 478  
 DB 419 LMASVQSKRRKRLRYLHCTNTDNPARYEGDLTLVAINLHVNTKYLRLPYFSNKOVDKYL 480  
 QY 421 LMASVQSKRRKRLRYLHCTNTDNPARYEGDLTLVAINLHVNTKYLRLPYFSNKOVDKYL 480  
 DB 421 LMASVQSKRRKRLRYLHCTNTDNPARYEGDLTLVAINLHVNTKYLRLPYFSNKOVDKYL 480  
 QY 479 LIRPLGPHGLLSKSVQNLGLTLKAVDDOTLPLMEKPLRPSSISGLPFAFSYFFVIRNAK 538  
 DB 479 LIRPLGPHGLLSKSVQNLGLTLKAVDDOTLPLMEKPLRPSSISGLPFAFSYFFVIRNAK 540  
 QY 481 LIRPLGPHGLLSKSVQNLGLTLKAVDDOTLPLMEKPLRPSSISGLPFAFSYFFVIRNAK 540  
 DB 481 LIRPLGPHGLLSKSVQNLGLTLKAVDDOTLPLMEKPLRPSSISGLPFAFSYFFVIRNAK 540  
 QY 539 VAACT 543  
 DB 539 VAACT 543  
 QY 541 VAACT 545  
 DB 541 VAACT 545

## RESULT 14

AAV34173  
 ID AAV34173 standard; Protein; 530 AA.

AC AAV34173;  
 DT 15-NOV-1999 (first entry)  
 XX  
 DE Human pre-proheparanase protein sequence.  
 XX  
 KW Human; pre-proheparanase; platelet; wound healing; angiogenesis blocker;  
 KW inflammation; psoriasis; diabetic retinopathy; solid tumor; arthritis;  
 KW heparin degradation; anticoagulant neutralisation; asthma; CNS disease;

inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;  
 tumor growth; fibroproliferative disorder; neurodegenerative disease;  
 therapy.

Homo sapiens.  
 WO9943830-A2.

02-SEP-1999.  
 18-FEB-1999; 99WO-US01489.

26-MAR-1998; 98US-0079401.  
 24-FEB-1998; 98US-0075706.

(PHAA ) PHARMACIA & UPJOHN CO.

Faibanks MB, Heinrikson RL, Mildner AM,  
 WPI; 1999-540598/45.  
 N-PSDB; AA211236.

New isolated platelet heparanase polypeptides, used to develop  
 products for, e.g. wound healing and blocking angiogenesis

Claim 12; Fig 7; 57pp; English.

This sequence is the human pre-proheparanase of the invention. This  
 sequence was isolated from human platelets. The heparanase can be used  
 for identifying agents which alter heparanase activity. The heparanase  
 can be used for wound healing or for blocking angiogenesis or  
 inflammation. It can be used for treating e.g. psoriasis, diabetic  
 retinopathy or solid tumors, or for the degradation of heparin and the  
 neutralisation of heparin's anticoagulant properties during surgery.  
 Inhibitors of heparanase activity can be used in the treatment of  
 arthritis, asthma, and other inflammatory diseases, vascular restenosis,  
 atherosclerosis, tumor growth and progression, fibroproliferative  
 disorders, and central nervous system (CNS) and neurodegenerative  
 diseases. The products can also be used for detection and diagnosis. The  
 purified heparanase, both recombinantly produced human heparanase and  
 heparanase isolated from human platelet activity, allows for the  
 convenient selection of compounds having anti-heparanase activity,  
 i.e. inhibitors of heparanase activity, by measuring inhibition of  
 heparanase activity. Inhibition of heparanase activity can be measured by  
 blocking heparanase-mediated release of radioactive fragments from in  
 vivo radiolabelled (HSPG)/heparin.

Sequence 530 AA;  
 Query Match 97.3%; Score 2764; DB 20; Length 530;  
 Best Local Similarity 99.4%; Pred. No. 5.4e-266;  
 Matches 527; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

14 MLLLGPPPLSPGALPPPAQADVDVLDLPTFOEPHLVSPGLSTYTDANATDPRFL 73  
 1 MLLLGPPPLSPGALPPPAQADVDVLDLPTFOEPHLVSPGLSTYTDANATDPRFL 60  
 74 LGSPPPLTARGLSPAYLRPGSTKMDPLIPPKKSTEBBSWMOSONODICKKSGSP 133  
 61 LGSPPPLTARGLSPAYLRPGSTKMDPLIPPKKSTEBBSWMOSONODICKKSGSP 120  
 134 PVEBKRLLEWPPQEOBLREHYOKKFKKSTYSRSSVDLYTFANCSGLDLFGNALMR 193  
 121 PVEBKRLLEWPPQEOBLREHYOKKFKKSTYSRSSVDLYTFANCSGLDLFGNALMR 180  
 194 TADLQNNSSNAQLLDYSSKGNISWEIENPNSLTKADIFINSQIGBYIQHLK 253  
 181 TADLQNNSSNAQLLDYSSKGNISWEIENPNSLTKADIFINSQIGBYIQHLK 240  
 254 RSTFPAALYGVDPGPRKTKAKMLKSPFKAGGEYIDSTWHYLYLNKRTATREDFLNP 313  
 241 RSTFPAALYGVDPGPRKTKAKMLKSPFKAGGEYIDSTWHYLYLNKRTATREDFLNP 300

314 DVLDIFISSVQKQVYVESTRPGKWLIGETSSAYGGGAPLSDTEAAGPMMLDKIGLSH  
 301 DVLDIFISSVQKQVYVESTRPGKWLIGETSSAYGGGAPLSDTEAAGPMMLDKIGLSH 360  
 374 RMGIEVVMRQVFPAGNHYLVENFDPDPYMLSLFLFKLVGRVMAASVQSKRRRLRY 433  
 361 RMGIEVVMRQVFPAGNHYLVENFDPDPYMLSLFLFKLVGRVMAASVQSKRRRLRY 420  
 434 YLHCTNTDNPYKGGDLTYAININAHNTKYLRLPYPSNKOVDKYLRLPGHGLLSKV 493  
 421 YLHCTNTDNPYKGGDLTYAININAHNTKYLRLPYPSNKOVDKYLRLPGHGLLSKV 480  
 494 QLNGLTLKMDVDTPLPLMEKPLRPGSSGLPAPSVSEFPYIRNAKVACI 543  
 481 QLNGLTLKMDVDTPLPLMEKPLRPGSSGLPAPSVSEFPYIRNAKVACI 530

RESULT 15  
 AA17083  
 ID AA17083 standard; Protein; 532 AA.  
 AC AA17083;  
 XX 21-JUL-1999 (first entry)  
 DT Seq ID No: 15 of WO9921975.  
 DE  
 XX Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;  
 KM metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;  
 KM atherosclerosis; atherosclerosis; inflammation; tissue development;  
 KM human; HSPG.  
 OS Homo sapiens.  
 XX WO9921975-A1.  
 PD 06-MAY-1999.  
 XX 28-OCT-1998; 98WO-AU00896.  
 PF 09-DEC-1997; 97AU-0000812.  
 PR 28-OCT-1997; 97AU-0000062.  
 XX (AUS ) UNIV AUSTRALIAN NAT.  
 PA Freeman CG, Hamdorf BJ, Hulett MD, Parish CR;  
 PI WPI; 1999-312956/26.  
 DR N-PSDB; AAX37260.  
 DR polynucleotides encoding mammalian endoglucuronidases, especially  
 heparanases, useful to promote wound healing

Claim 6; Page 76-79; 112pp; English.

The invention relates to nucleic acid sequences that encode heparanase  
 enzymes having endoglucuronidase activity. Recombinant heparanases are  
 capable of removing the HS side chain from heparan sulfate proteoglycan  
 (HSPG). Sulfated oligosaccharides, sulfonates or HSPG can be used to  
 inhibit heparanase, this is useful for treatment of a physiological or  
 medical condition associated with elevated heparanase activity, such as  
 metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,  
 atherosclerosis, atherosclerosis and inflammation. The human, murine and  
 rat heparanases can be used to enhance wound healing, especially  
 associated with tissue development and repair. The conditions mentioned  
 above can be diagnosed using specific antibodies, and also using primers  
 and probes specific for the heparanase polynucleotides. Other uses of the  
 heparanases include sequencing sulfated molecules such as HSPG.

Sequence 532 AA;  
 Query Match 96.3%; Score 2737; DB 20; Length 532;  
 Best Local Similarity 99.8%; Pred. No. 2.7e-263;

Matches 522; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MLRSFGALPPPLMLLLPLGLSPGALPRPAQADVVLDLFTQEPHLVSPSLSYT	60
Db	1	MLRSFGALPPPLMLLLPLGLSPGALPRPAQADVVLDLFTQEPHLVSPSLSYT	60
Qy	61	IPANLATDRPFLILGSPKLTARGLSPAYLRFQGTKDPLIPDPKSTFEERSYQS	120
Db	61	IPANLATDRPFLILGSPKLTARGLSPAYLRFQGTKDPLIPDPKSTFEERSYQS	120
Qy	121	QVNOODICKYGIIPDVEEKLRLWPYQEODLLREHYQKFKNSTYSRSSVDLYTFANCS	180
Db	121	QVNOODICKYGIIPDVEEKLRLWPYQEODLLREHYQKFKNSTYSRSSVDLYTFANCS	180
Qy	181	GLDLIFGLNALRTADLQWSSNAQLLDYCSSKGYNISWEIGNEBNSFLKADIFINGS	240
Db	181	GLDLIFGLNALRTADLQWSSNAQLLDYCSSKGYNISWEIGNEBNSFLKADIFINGS	240
Qy	241	QEGEDYIOLHKLKSTFKNAKLYGPDVGOPRRKTAQMLKSFLLKAGGEVIDSVTHHYYL	300
Db	241	QEGEDYIOLHKLKSTFKNAKLYGPDVGOPRRKTAQMLKSFLLKAGGEVIDSVTHHYYL	300
Qy	301	NGRTATREDFLPNDVDLFISSVQKVFQVVESTRPGKVMIGETSSAYGGAPLLSDTFA	360
Db	301	NGRTATREDFLPNDVDLFISSVQKVFQVVESTRPGKVMIGETSSAYGGAPLLSDTFA	360
Qy	361	AGFMWLDKLGLSARNGIEVVMRQVFFGAGNTHLVENEDPLPDYMLSLFKKLVGTVLM	420
Db	361	AGFMWLDKLGLSARNGIEVVMRQVFFGAGNTHLVENEDPLPDYMLSLFKKLVGTVLM	420
Qy	421	ASVQSKRRRLRYLHCTNDNRYKEGDLTYAIVLHNVTKYLRPYPEFSNKQYDKYLL	480
Db	421	ASVQSKRRRLRYLHCTNDNRYKEGDLTYAIVLHNVTKYLRPYPEFSNKQYDKYLL	480
Qy	481	RPLGPHGLSKSVQVNLGTLKAVDDOTLPLMEKPLRPGSSLG	523
Db	481	RPLGPHGLSKSVQVNLGTLKAVDDOTLPLMEKPLRPGSSLG	523

Search completed: October 22, 2003, 20:24:22  
Job time : 88 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using SW model

Run on: October 22, 2003, 20:22:55; Search time 42 Seconds

(without alignments)  
1243.324 Million cell updates/sec

Title: US-09-759-207-2

Perfect score: 2842

Sequence: 1 MLRSKPALPPMLMLLGP.....LPASYSFFVINAVACI 543

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : PIR 76:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	897.5	31.6	480 2	JC7506
2	416	14.6	521 2	T45608
3	169.5	6.0	190 2	T01953
4	112.5	4.0	356 2	P64383
5	111.5	3.9	575 2	T12094
6	111	3.9	670 2	T10666
7	111	3.9	688 2	S32961
8	111	3.9	2298 2	T45648
9	109.5	3.9	879 2	E91031
10	108.5	3.8	411 2	S74760
11	107.5	3.8	500 2	D87541
12	106	3.7	879 2	P85875
13	105	3.7	670 2	T38446
14	104.5	3.7	788 1	S00652
15	104	3.7	432 2	P70411
16	104	3.7	2013 2	A11489
17	103.5	3.6	587 2	S36231
18	103.5	3.6	676 2	AF1153
19	103.5	3.6	687 2	P85188
20	103	3.6	796 2	D97065
21	101	3.6	594 2	A82913
22	101	3.6	644 2	A97268
23	100.5	3.5	805 2	C86525
24	100.5	3.5	805 2	H72098
25	99.5	3.5	989 2	AB2140
26	99.5	3.5	510 2	H69893
27	99.5	3.5	837 1	A31842
28	99	3.5	897 2	G02529
29	99	3.5	4644 1	A38905

30	98.5	3.5	596 2	T04506	hypothetical prote
31	98.5	3.5	629 2	C64180	hypothetical prote
32	98.5	3.5	654 2	T14202	NADH2 dehydrogenas
33	98.5	3.5	699 2	P95146	DNA topoisomerase
34	98.5	3.5	701 2	D98014	DNA topoisomerase
35	98.5	3.5	746 2	T46821	sideophore recept
36	98.5	3.5	746 2	A85420	Rhizobactin r
37	98.5	3.5	1012 2	UC5925	membrane K10cho pr
38	98	3.4	465 2	T19113	hypothetical prote
39	98	3.4	716 1	C60008	RNA-directed RNA p
40	98	3.4	760 2	T34414	hypothetical prote
41	98	3.4	817 2	H75035	probable membrane
42	97.5	3.4	454 2	T20829	probable serine ca
43	97.5	3.4	511 2	S61166	probable membrane
44	97.5	3.4	604 2	E75119	hypothetical prote
45	97.5	3.4	804 2	G71546	probable DNA gyrase

## ALIGNMENTS

### RESULT 1

JC7506

heparanase protein 2a - human

C/Species: Homo sapiens (man)

C/Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 01-Dec-2000

C/Accession: JC7506

R/McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hancock, M.; Pe

Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000

A/Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family m

A/Reference number: JC7506

A/Accession: JC7506

A/Molecule type: mRNA

A/Residues: 1-480 <MCR>

A/Cross-references: GB:AF282885

C/Comment: This protein, a intracellular membrane-bound enzyme, has biological and ther

C/Genetics:

A/Gene: hpa2a

A/Map position: 10q23-10q24

C/Keywords: heparin binding; membrane bound

Query Match	Score	DB 2	Length
Best Local Similarity	36.0%	Pred. No. 2.8e-59;	
Matches 202; Conservative 74; Mismatches 146; Indels 139; Gaps 9;			
Query Match	31.6%	Score 897.5;	DB 2; Length 480;
20 PLGPIPGAL-----PRPA-----QAQDVLDPTQEPHLVPS 55			
18 PPAIAPALYIALLLSLSSQAGDRPLVDRAAGLKEKTLILLDSTKNPRTVNN 77			
56 FLVVTIDANLTDPRFLLGSPKRLTARGLSPAYRFGGKTDFLF---DPKKEST 111			
78 FLISGLDLSIHD-GMDLFLSKRLVTLAGSPFAFGKRTDFLFQULRNPAKR- 135			
112 FEERSTWOSQVNDICCKGSIIPDVEEKLRLWPYQEBLLREHYQKFKNSTYRSRSD 171			
136-----GCGPDP-----YLNKYE----- 148			
172 VLYTFANGSGDLIFGNALRLTADLQWSSNAQLLDLYCSSKGVNIMWELGNEPNSFLK 231			
149-----DEPNRT 156			
232 RADIFINGSGEDYIQAHLKLR-STFKNAKLYGPDVGQPRRTAKMLKAGKEVI 290			
157 MGRVAVNSQKQKYOIKSLQPIRISRASISGPNIGRRKRVIALLDGPMKAVGSTV 216			
291 DSVTTHHTYLANGRTREDFLNPVDLFISSVQKVPQVVESTPRGKVMIGFTSSAYG 350			
217 DAVTWQHCYIDGRVVKWDFPKLRLDLDQIRIQCVVTVYFGKKIMEGVVTSAG 276			
351 GAPLISDFAGFMWLDLGSARNGIEVVRQVFFGAGNHLVDENPDLPDYLSLIF 410			
277 GTNNLSDSYAGFMWNLGLANOGIDIVIRHSFFDHGYNHLVDQNPPLPDYLSLIF 336			

QY 411 KKLVTGKVLMAASVQSGSKRR-----KLRVYLHCTNTDNPYKKGEDLTLYAINLHANT 461  
 DB 337 KKLIGKVLAVHAGVAGQKRRPGRVIRDRKIRYAHCTNNHNYVAGSITLFLINLHNSR 396  
 QY 462 KTLRLPYPSNQVDKYLRLPDPHGLSKSVQVNLTLKMDVDTLPLMEKPLRPSSG 521  
 DB 397 KKLKLAGTLRDLKLVHGYLLQPYGOELKSKSVQVNLQPLVMDVDTLPLMEKPLRPAGRT 456  
 QY 522 LGLPAPSYSPFYIRNAKVAAC 542  
 DB 457 LVIPVYMGPFVKNVNALAC 477

RESULT 2  
 T45608  
 Hypothetical protein F13G24.30 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C:Accession: T45608  
 R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voel, M.; Robben, J.; Volckaert, G.; Be  
 submitted to the Protein Sequence Database, December 1999  
 A:Reference number: Z23009  
 A:Accession: T45608  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-521 <BBV>  
 A:Cross-references: EMBL:AL133421  
 A:Experimental source: cultivar Columbia; BAC clone F13G24  
 C:Genetics:  
 A:Map position: 5  
 A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3  
 A>Note: F13G24.30

Query Match 14.6%; Score 416; DB 2; Length 521;  
 Best Local Similarity 29.2%; Pred. No. 3.5e-23;  
 Matches 154; Conservative 68; Mismatches 184; Indels 122; Gaps 24;

QY 75 LGSPLRLTARGLSPAYLRFSGTKDPLFDPKKESTFEERSYQWQVNDLCKXGSIIP 134  
 DB 55 LTRPLTLTAIKAPKPLRIRIGSLQDQVLYDVGNLKT-----PCR----- 94  
 QY 135 DVEBKRLLEMPYQOQLLREHYOKFKNS---TYSRSV---DLYTFANCGLDLIF 186  
 DB 95 -----PQKMK-----NSGLFGFSKGLHMRMDELNSFLRATGAVTF 132  
 QY 187 GLNALRLTADLQ-----WSSNAQLLDYSSKGYNI-SWELGNEPSEFKKADIFIN 238  
 DB 133 GLNALRGRHKLKAGKMGAMDHINQDPLNTYTSKGYVIDSWFEGNELSG--SGVCASVS 190  
 QY 239 GSQLGSDYQLHKLRLKSTFKNAKXGPRVGP-----RKRAKMLKSLKAGSEVDSV 293  
 DB 191 AELYGDLVLVDKVIN-KYKSMWLKPLVAPGGFYEQWYTKLEI---SGPSVADV 246  
 QY 294 TWHHYLYNGRT--ATREDFLNDVDLDFISSYQKVF---QVVESTPGKRWLGETSSA 347  
 DB 247 THHITNLGSGNPLVYKIMDS---YLSQSKTFKVNQIIOHGFAPASFWGESGA 302  
 QY 348 YCGAAPLSDTTPAAGFMWLKGLSLKMGIEVVMQVFFGAGNYHVLDE-NFDPPLDYWL 406  
 DB 303 YNSGGRHNSDTFIDSEFWYLDQGLMSRHNHTKYCQTLVG-GFYGLLEKGTVPMPDDYS 361  
 QY 407 SLTFKKLVOTKYLMAASVQSGSKRKRLRYVLAHCNTNDPRKEDDLTYAINLHANTKYL-- 464  
 DB 362 ALMLHRLMGKGLAVVOTDGP--QLRVYAHCSK-----GRAGVTLTLNLISNODPVS 413  
 QY 465 -----RLPYPS--NKQVDKYLRLP--LGPHG--LTSKSVOL 495  
 DB 444 VENGINVLVMAASRKKKSLDLTKRPFWSIGSKADGYLNRREHYLTPENGVLARKTMYL 473  
 QY 496 NGTLTKMDVDTLPLMEKPLRP-GSSGLPAPSYSPFYIRNAKVAAC 542  
 DB 474 NKSLSKPLRTATGDIPIST-EFVLRSVNSPLNVLPLMSFIVLPNFDASAC 520

RESULT 3  
 T01953  
 Hypothetical protein T2L5.6 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 21-Jan-2000  
 C:Accession: T01953  
 R:Geisels, C.; Smith, A.; Le, T.  
 submitted to the EMBL Data Library, October 1998  
 A:Description: The sequence of A. thaliana T2L5.  
 A:Reference number: Z14470  
 A:Accession: T01953  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-190 <GBL>  
 A:Cross-references: EMBL:AF096371; NID:g3695386; PID:g3695392  
 A:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 36/2; 69/3  
 A>Note: T2L5.6  
 C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Query Match 6.0%; Score 169.5; DB 2; Length 190;  
 Best Local Similarity 27.8%; Pred. No. 2.1e-05;  
 Matches 54; Conservative 34; Mismatches 57; Indels 49; Gaps 9;

QY 382 ROVFFGAGNYHVLVD-ENFDPPLDYWLSLFKKLVGTQVLMASVQSGSKRKRLVYLHCTNT 440  
 DB 12 RQSLIG-GNYGLNTVTFETPNPDYASALIRQLMGKRALFTTFSGTK--KIRSYTHCA-- 66  
 QY 441 DNPYKKGEDLTLYAINLHNV-----TKYRLRLPYPSNQVDKYLRLP 483  
 DB 67 ---RQSKG-ITVLLMNDLNTTIVVAKVYELNNSFSLRHTKHK-----SYKASSQLFG-- 115  
 QY 484 GPHGL-----SKSVQLNGTLTKKMDVDTLPLMEKPLRPSSGLPAPFS 528  
 DB 116 GPNGVIOGEBYHLAKQONLHSGTMLNGLNGLQVNSMGDPLPIPIHNSTEPITIAYS 175  
 QY 529 YSPFYIRNAKVAAC 542  
 DB 176 IVFVHMRVAVVAPAC 189

RESULT 4  
 F64383  
 Hypothetical protein MJO670 - Methanococcus jannaschii  
 C:Species: Methanococcus jannaschii  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: F64383  
 R:Bult, C.V.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
 Reich, C.I.; Overbeek, R.; Kirschner, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;  
 Iron, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.P.; Roberts, K.M.; Hurest, M.A.  
 Science 273, 1058-1072, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C  
 Article: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
 A:Reference number: A64300; MUID:96337999; PMID:8688087  
 A:Accession: F64383  
 A:Molecule type: DNA  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Residues: 1-356 <BLU>  
 A:Cross-references: GB:U67514; GB:L77117; NID:g2826304; PIDN:AA898664.1; PID:g1591384; 1  
 A:Map position: REV596956-595886  
 A:Start codon: GTG

Query Match 4.0%; Score 112.5; DB 2; Length 356;  
 Best Local Similarity 21.2%; Pred. No. 1;  
 Matches 85; Conservative 48; Mismatches 143; Indels 125; Gaps 18;

QY 126 ICKT-----GSIPDVEKRLLEMPYQOQLLREHYOKFKNSYSSVD----- 171

Db 14 IRKTKYNGKREKIKERLKE-----LKEHVLYVETEDGTYTLAEDDEEMMSKV 66  
Qy 172 -----VLYTFANCGLDILFGLMALLRTADLOMSSNQLLDYSSKGYINISWELGNBP 226  
Db 67 GALKKAIYKFAKPS-----KITDL-----SNPR-VLDLCSGMGYNAIMALHYNK 109  
Qy 227 NS-----FL-----KKADITINGSLGSDYIQLHKLKSTF 258  
Db 110 NAEIDMVEICSEVLEPLDIPYKEHEIIMDKVYEFLN-KIGIEY-----KSDY 159  
Qy 259 KNAKLYGPDVCGPRKTKAMLSPLKAGEVIDSYTMHHYLYNGRTA--REDPLNPVL 316  
Db 160 DNINLY--VGDARFIIKSDKY-----NVFHDARSPPKDPFLTYTDFL----- 202  
Qy 317 DIFISSVQKRVQVVESTPRGKKVWLGETSAYAGGAPLLSDTFAAGFMWLDLGLSABMG 376  
Db 203 -----KEIKRMBDN--GVLI-----SVSSALPRSMALVDCGFYISEKESVGRKRG 246  
Qy 377 IEVWROVFPFAGNHYLVNDENP-----PLPDVMSLLEPKLVGTVMASVQSGSR 429  
Db 247 ITLVYKKNFKNRINEVDERVIATLALPYRDETLTLTKDKIIEDEEREERKLEKLI 306  
Qy 430 KLRVYIHCNTNDNPRYKSGDLTVA--INLHVTKYLRPY 468  
Db 307 KIGKTLSTKQIKKGNIPBEILKIQKEDLANSSEIIRKMLKF 347  
RESULT 5  
T12094  
beta-fructofuranosidase (BC 3.2.1.26) - fava bean  
C/Species: Vicia faba (fava bean)  
C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C/Accession: T12094  
R/Meber, H.; Borstjuk, L.; Heim, U.; Buchner, P.; Mobus, U.  
Plant Cell 7, 1835-1846, 1995  
A/Title: Seed coat-associated invertases of Fava bean control both unloading and storage  
A/Reference number: Z17416; MUID:96093423; PMID:8535137  
A/Accession: T12094  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Structure: preliminary; translated from GB/EMBL/DBJ  
A/Residues: 1-575 <WEB>  
A/Molecule type: mRNA  
A/References: EMBL:Z35162; NID:g861154; PION:CAA6526.1; PID:g861155  
A/Experimental source: cv. Fido, seed coat  
C/Genetics:  
A/Gene: CWINV1  
C/Superfamily: beta-fructofuranosidase  
C/Keywords: cell wall; glycoprotein; glycosidase; hydrolase  
Query Match 3.9%; Score 111.5; DB 2; Length 575;  
Best Local Similarity 21.4%; Pred. No. 2.5;  
Matches 72; Conservative 48; Mismatches 107; Indels 109; Gaps 19;  
Qy 46 QEPHLVLS-----PSFLSVITIDANLTPRFLILLSPKRLTLARGLS-----P 89  
Db 228 KPHISAKRTGMBCPDFPVSLBSKNGLD--LSMMGNVVKHVLKNSLDITREYTYTG 285  
Qy 90 AYLR-----FGGKTDF-----LIFPKKSTFEERSYV--OSQVNO 124  
Db 286 TYLQNDKTYIDPKISEDMGGLRYVGNFYASKSFDPDK--NRIIWMANSDRKE 341  
Qy 125 DICRYG--SIPDV-----EEKLRLEMPYQEOQLLR---EHYQKPKNSTYSRSRV 170  
Db 342 DDVKMGAGIQAIPRTWLDSSRRQLR-QMPVEELNRLRGQVEKMKRLKKGYG--L 396  
Qy 171 DVLTFPANCGLDLFGNALILRTADLOMSSNQLLDYSSKGYINISWELGNPNFL 230  
Db 397 EVKGITAGQADVEVFSFSLDKAEPDPMENAR--DLCAQKSKVRGGVG--PGLL 451  
Qy 231 KKADIFINGSLGSDYIQL-----HKL-----RKSTFKNAKLYGP-----DV 268  
Db 452 TLA-----SKLEBYTSVFFRVRFKANHKILMCSDAKSSSLNLEIVKPSFAGFVNDL 505  
Qy 269 GQPRKRTAKMLKSLKAGEVIDSYTMHHYLYNGRT 304

Db 506 GNNKLSLSL-----IDHSVVSFGVGCKT 531  
RESULT 6  
T10666  
hypotheical protein F6E21.40 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Feb-2001  
C/Accession: T10666  
R/Beyan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro  
submitted to the Protein Sequence Database, June 1999  
A/Reference number: Z16533  
A/Accession: T10666  
A/Molecule type: DNA  
A/Residues: 1-670 <BEV>  
A/Cross-references: EMBL:AL049914; GSPDB:GN00062; ATSP:F6E21.40  
A/Experimental source: Cultivar Columbia; BAC clone F6E21  
C/Genetics:  
A/Gene: ATSP:F6E21.40  
A/Map position: 4  
A/Intons: 47/3; 87/1; 123/3; 203/3; 230/2; 255/3; 284/3; 305/1; 335/3; 347/3; 370/3; 35  
C/Superfamily: Schizosaccharomyces pombe negative regulator of mitosis skbl  
Query Match 3.9%; Score 111; DB 2; Length 670;  
Best Local Similarity 22.4%; Pred. No. 3.4;  
Matches 123; Conservative 77; Mismatches 194; Indels 156; Gaps 33;  
Qy 51 LVSPFLSVITIDANLTPRFLILLSPKRLTLARGSPAY--LRFGKTDFLIFPK- 107  
Db 47 LVDSYSPSLVEGN-GVDTVLPVCGSDLV-----LSPQMSHVVQKISWIDLDSGD 99  
Qy 108 -----KSTFEERSVQSOVNOIDICKYGIIPVBEKRLRMPYQQLIREHYQKKF 161  
Db 100 EVLMDSETTLKQIAATLHLSQMCB-----PD-----LTRPHYLAGL 139  
Qy 162 NSTYSRSV--DVLY-----FRANC--GLDILFGLMALLRTADLOMSSNQL 207  
Db 140 RVSCRSRSPISDEFELVKITFNQALTFGSSSLFCLNVIASALKMLRVPVLV--SEGD 196  
Qy 208 LDYSSKGYINISWELGN-----EPNSFLKKA-DIFYN-----GSOLGEDYIQLHKL 253  
Db 197 DD--TSEGLANDSWELNMSFRLCHDSKLVADLVLTLPSEISLGWMSGVS-VRAILS 253  
Qy 254 RKSTFKNAKLYGPDVGP--RRKTAUKSFL-KAGEVIDSYTMHHYLYNGRTAREDF 310  
Db 254 TDAFLTNAR-----GVPCLSKRHQKLIAGFPDHAQVVICGKRVHNLQKPLDSSSGTE 307  
Qy 311 LNPVLDIFISSVQKRVQVVESTPRGKKVWLGETSAYAGGAPLLSDTFAAGFMWLDL 370  
Db 308 KNP--LRIYLDYVAYVLPQKMSLSSEORIBLGYRDPLOAPLOPLMDNLEAQVETFE-- 362  
Qy 371 LSAWGIENVWROVFPFAGNHYLVNDENPDLPDVMSLLEPKLVGTVMASVQSGSR 429  
Db 363 --RDSVYIYOQ--RAVEKALVDR--VPDKASBL-----TYLVWVVGAGRGVL 406  
Qy 421 ASVQSGKR--RKLRYVICHNTNDNPRYKSGDLTVAINLHVTK-----YLRLEY 468  
Db 407 RASIQAAEEDTRKLVY--AVEKNPN-----AVVTLNVLKMGWEDVVTILISCDM 455  
Qy 469 PFSN--KQVDKYLRLPYGPHGLSKSVQVNLGLTKAMDQDTLPFLM--EKLAPGSLG 523  
Db 456 RFWVAPQADITLVSELGSRG-----DNLSPCLDGAQRFLEK--DGIS 498  
Qy 524 LPARSFV 533  
Db 499 IFS-SYTSFI 507  
RESULT 7  
S32961  
hypotheical protein YBR259w - Yeast (Saccharomyces cerevisiae)  
N/Alternate names: hypotheical protein YBR1727



C.Species: *Saccharomyces cerevisiae*  
 C.Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-Apr-2002  
 C.Accession: S33961; S66140  
 R.Doisignon, F.; Bileau, N.; Crouzet, M.; Aigle, M.  
 Yeast 9, 189-199, 1993  
 A.Title: The complete sequence of a 19,482 bp segment located on the right arm of chromo-  
 A.Reference number: S23346; MUID:93220397; PMID:8465606  
 A.Accession: S33961  
 A.Status: translation not shown  
 A.Molecule type: DNA  
 A.Residues: 1-688 <DOI>  
 A.Cross-references: EMBL:X70529; NID:91907246; PIDN:CAA49923.1; PID:9296558  
 R.Aigle, M.; Bileau, N.C.; Barthe, C.; Bileau, N.; Crouzet, M.; Doisignon, F.  
 submitted to the Protein Sequence Database, August 1994  
 A.Reference number: S45940  
 A.Accession: S46140  
 A.Molecule type: DNA  
 A.Residues: 1-688 <AIG>  
 A.Cross-references: EMBL:X36128; NID:9536684; PIDN:CAA85222.1; PID:9536685; MIPS:YBR259w  
 C.Genetics:  
 A.Cross-references: SGD:S0000463  
 A.Map position: 2R  
 C.Superfamily: *Saccharomyces cerevisiae* hypothetical protein YBR259w

Query Match 3.9%; Score 111; DB 2; Length 688;  
 Best Local Similarity 22.5%; Pred. No. 3.6;  
 Matches 67; Conservative 45; Mismatches 94; Indels 92; Gaps 16;

QY 126 ICKYGIIPDVEEKLLEMPYQOQLLREHYOKKFNSTYSRS-----VDVLYT 175  
 DB 164 MAEYSMSKWDSDKQQLQGMVEFRMKLKECLVKEFNFLQKSDPLKELIIPMEKIYV 223  
 QY 176 FANCGGLDIPGLNMLRTADLQWNSN-----AQLLD-----YSSSGY----- 216  
 DB 224 -NMC--TDFTBQVRIIDAEILMTSKNLFSSISSAVALRLNDLQMFSAFPPYGEALV 280  
 QY 217 -----NISWEIGNEPNSFLKKA--DIF--INGSQLG--BDYVQLHLKLR----- 255  
 DB 281 QDPFHRSLKMSNDKVESLIRALIFNDKFFPKNEQVUTKADGIFPLRLKRNKEHN 340  
 QY 256 -----STFKN--AKLYGPDVGQPRKTKAKMLKSLKAGGEV-----IDSV 293  
 DB 341 DVKDFHIQVIKYLNSQFKNNYSTLTMTSSKTQDRKSHNPPSSILDDGNKIGMHWSPIDE- 399  
 QY 294 TMAHYTLG-----RTATREDPLNPVDLDFISSQVQVQVVESTR--FGK 338  
 DB 400 -YSHFIDNDPLMRDKVYPKIYTNQTPPDASALFDS--HKIYAILISLRYLPEKR 454

## RESULT 8

T49648  
 Hypothetical protein B8B20.20 [imported] - *Neurospora crassa*  
 C.Species: *Neurospora crassa*  
 C.Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
 C.Accession: T49648  
 R.Schulze, U.; Aign, V.; Hohenseel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
 submitted to the Protein Sequence Database, May 2000  
 A.Reference number: Z25022  
 A.Accession: T49648  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-2298 <SCH>  
 A.Cross-references: EMBL:AL355933; GSPDB:GN00116; NCSP:B8B20.20  
 A.Experimental source: BAC clone B8B20; strain OR74A  
 C.Genetics:  
 A.Gene: NCSP:B8B20.20  
 A.Map position: 6  
 A.Introns: 426/3

Query Match 3.9%; Score 111; DB 2; Length 2298;  
 Best Local Similarity 19.3%; Pred. No. 22;  
 Matches 114; Conservative 79; Mismatches 190; Indels 208; Gaps 28;

QY 78 PKRLTLARGLSPA-----YLRFGTKTDFLIPDPKSTEEFRSYQOV-NODIC 127  
 DB 1447 PRVDLIERLITPSNAHKEACILINIRANNOALRVVSGBSASRPITRNANVFNILD 1506  
 QY 128 KYGSIIPDVEEKL-----LEWYQOQLREHYOKKFNSTYSRSVDVLYTPAN 178  
 DB 1507 QYMAESDIEQFPALGAENNRSIDAAMRELLTKN-----KAVALDILHTSAR 1555  
 QY 179 CSGDLIFGNAL-----LRTDILQ-----WNSNAQLLDYSSKGYNI 218  
 DB 1556 AS-LDVAKQAKTEALAYTLAVTQLOKCTTLHFGSGFPDGLINVAL-----DTHAFL 1609  
 QY 219 SW-ELGHEPNSFLKADIFINGSQLGHDYQLHLKRSFKNAKLYGPDVGQPRKRTAK 277  
 DB 1610 CMITSSEBOYSSSESSADIDPRQEDAILLLQELKKEPFMA-----RELLAL 1659  
 QY 278 MLKSFPLAGSEVID-SVYWHYVYNGTATREDPLNDVDLDFISSQVQVQVVESTRPG 336  
 DB 1660 PLKATTFGKQTEQVACTEKTVTTLAALTAAR-----FIO--ERYVQVLPYEQPG 1706  
 QY 337 K-----KWLGETSSAYGGAPILSDTPAG-----FMWLDK 368  
 DB 1707 KYGLPDPMPKMSGPERRWL-----PLFIATLVNNKVPDFKDIETNLSLWQS 1755  
 QY 369 LGLSAR-MGLEVVRQVFPFGAGNYHL--VDENFDPLDPYMLSL-LPKLVG--TKVL-- 419  
 DB 1756 IIKMRPLGTYTTLAEVLQGRGLPFLAEADVSAGMTDYNHDLFSRAIHYMKALRG 1815  
 QY 420 -----NASVQSK-----RRKLRYLHCTNTDPRYKESDGLTYVAINLHNTKYL 464  
 DB 1816 ATPAPGVTSASAYGSSASQSIROREBFSH----- 1847  
 QY 465 RLPPFSNKQVDKYLPLIG-----PHGLSKSVQNLGLTKMD-- 504  
 DB 1848 TLQIAMTNIKDLDLNSLALADPTASTERHRYMAFTNCLIS-LKSHGVGVVVDSE 1906  
 QY 505 -----DQTLPLMEKPLRPG-----SGLGLPAVSYSF-FVIRNAKVA 540  
 DB 1907 FLTSDSVSPPLQDPQLHTAGIMAYGVRLSKQDVPASQLEWYLFNNPKVA 1957

## RESULT 9

E91031  
 Probable outer membrane protein Ec93221 [imported] - *Escherichia coli* (strain O157:H7, e  
 C.Species: *Escherichia coli*  
 C.Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C.Accession: E91031  
 R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A.Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genc  
 A.Reference number: A96429; MUID:21156231; PMID:11258796  
 A.Accession: E91031  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-879 <HAY>  
 A.Cross-references: GB:BA000007; PIDN:BA036644.1; PID:913362691; GSPDB:GN00154  
 A.Experimental source: strain O157:H7; substrain R1MD 0509952  
 C.Genetics:  
 A.Gene: Ec93221

Query Match 3.9%; Score 109.5; DB 2; Length 879;  
 Best Local Similarity 20.1%; Pred. No. 6.7;  
 Matches 130; Conservative 68; Mismatches 208; Indels 241; Gaps 33;

QY 52 VSPSEFLSTYIDANLATPRLPILLGSPKRLTLARGLSPAYLFGQTKTDFLIPDPKEST 111  
 DB 20 MSGSYVMAAENBIQPDSEFLKGDFTKI-DLKPFSSGCVYEPG--KYNLQVQNKQPLT 76  
 QY 112 FEENSYQOVNODICKYGIIPDVEEKL-----RLEWYQOQLREHYOKKFNSTYS 163  
 DB 77 EBYDIWYVASNDASKTYVACTPBLVAQFLKEDVARNLQWIDHGKCLKKEQLE----- 130

QY 244 TYSRSSVDVLYTPANCGLDLIFGH-NALLRTADLQNNSSN-----AQLLLDYC----- 211  
 DB 131 GIDIK-ADLSGALVLSLPOAVLEETDINMPSPKMDGICGLADYSITNOT 182  
 QY 212 -----SSKGINI-SMEL-GNEPNSFL-KKADIFINGSO----- 241  
 DB 183 RHEKGGDDSNISGNQGVGMGLMRRLADQDTYLSKSNDDVDVINGDDTQKMEWSR 242  
 QY 242 -----LGRDYIQLHLLKSTF-----KKAKLYGGDV 268  
 DB 243 YYAMFALPSLAKKLDLGEDY-----LNSDIFDGENYVGGISITDDQMLPPLRLKGYADI 296  
 QY 269 GQPRKRTAKMLKFLKAGEVY-----DSVTMHHYLYNGRTATREDFLN 312  
 DB 297 SGVATTTAKVYVSOQ-GRVYEVQVAPGPRRIODLGDV-----SGTLHRIERON 346  
 QY 313 PDVLDFISSVQKVPQVVESTRPK-KVWL-----GETS----- 345  
 DB 347 GVOVEYDINTASMP-----LTRPGQVRYKLMGRPOEWGHHVEGGFSGGEGASMGIANGM 402  
 QY 346 SAYGGAGLLSD-----TFNA-----GFMWLDL-----GLSAR 374  
 DB 403 SLVYGA-----LDEHYQSAALGVGRDLSVFGAFAFDITSHTRLDKETAYGKGLDGNFR 459  
 QY 375 MGI-----EVMWRQVFPAGVYHLVDENFDPLDYLILFKLVGT-----KVLASVQGS 426  
 DB 460 LSYSDPPELMSRYTFAG-YSRESENFMTMSE-LIASSEWRTDNDKMTATATNO 515  
 QY 427 KRRKRLVYLCTNDNPRYKEDLTLVAI-----NLNNTK----- 462  
 DB 516 NFRDQGVVLYNLTHTYWDREQTNNVNLSHYNGLSIRMSISMGRYREYDNOAK 575  
 QY 463 -YLRPPSPSKQVDKYLRLPRLGHLKSLKSVQNLGLTKMVDQT 507  
 DB 576 GVYISLSPMGDSSSTISY-----NGMYGSGSDSSQVG-YFSKVDAT 616

## RESULT 10

S74760  
 hypothetical protein alr1617 - *Synechocystis* sp. (strain PCC 6803)  
 C/Species: *Synechocystis* sp.  
 A/Variety: PCC 6803  
 C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
 C/Accession: S74760  
 R/Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
 DNA Res. 3, 109-136, 1996  
 A/Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
 sp.  
 A/Reference number: S74322; MUID:97061201; PMID:8905231  
 A/Accession: S74760  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-411 <KAN>  
 A/Cross-references: EMBL:D90901; GB:AB001339; NID:g1651897; PTDN:BA16911.1; PTD:dl01764  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 3.8%; Score 108.5; DB 2; Length 411;  
 Best Local Similarity 21.4%; Pred. No. 2.5;  
 Matches 66; Conservative 55; Mismatches 94; Indels 93; Gaps 17;  
 QY 159 KFKNSTYRSVDVLYTPANCGLDLIF--GLNALLRTADLQNNSSNAQL----- 206  
 DB 151 EFRILSPREIDIT--FAGSTKLDLLASBENIDCIYHLANPRVYTSVAVMGQTLTMLRN 207  
 QY 207 LLDYSSKSG--YNIWMEL-----GNEPNSFLK-----ADIFINGSQGE 244  
 DB 208 VIDGLADIDPLIYSSWEIYSYGAGTTHADBSTALRGPEYKTYLAELI----- 260  
 QY 245 DYIQLHLKLRSTFNALYGPVQGPRTAKMLKSLKAGEVIDSVTWHHYLYNGRT 304  
 DB 261 DHCRTREGLRCAILRSSPVYSGMSDKP-----KFIENFFKASQOKQIVT--HHYING-- 311

QY 305 ATREDPLNDV-----LDIFISSVQKVPQVVESTRPKKWLGETSSAYGAGAPLLSDTFA 360  
 DB 312 -----NPKLDLHIDDLISSIVATL-----KSRIGINNT-----CTGLSSTLK 351  
 QY 361 AGFWMLDKLGLSA-----RKGIEVWRQVFPAGVYHLVDENFDPLDYLILFKLVG 415  
 DB 352 LAEMIRDELSSSMITQOIEVYTEVASIANNVGRAN-HVLD-----WEPVIFB-QG 400  
 QY 416 TKVLMASV 423  
 DB 401 LKSLHQI 408

## RESULT 11

D87541  
 beta-xylosidase [imported] - *Caulobacter crescentus*  
 C/Species: *Caulobacter crescentus*  
 C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C/Accession: D87541  
 R.Nieman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Bisen, J.; Heidelberg, J.  
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.O.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolor  
 n, J.; Rmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A/Title: Complete Genome Sequence of *Caulobacter crescentus*.  
 A/Reference number: A87249; MUID:21173698; PMID:11259647  
 A/Accession: D87541  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-500 <STO>  
 A/Cross-references: GB:AE005673; NID:g13423886; PTDN:AAK24328.1; GSPDB:GN00148  
 C/Genetics:  
 A/Gene: CC2357

Query Match 3.8%; Score 107.5; DB 2; Length 500;  
 Best Local Similarity 25.7%; Pred. No. 4;  
 Matches 56; Conservative 35; Mismatches 80; Indels 47; Gaps 15;

QY 165 YSRSSVDVLYTPANCGLDLIFGN--ALLRTAD--LQW--NSSNAQL-----LID-- 209  
 DB 81 YDWTKIDQLYDALLAKGKIFELGFTPEAKMTSDQITFYWKGTSHPKLGPMRDLDIAF 140  
 QY 210 -YCSSKSGYNI--SW--ELGNEBN--SPKKADIFINGSQLGEDYQLHLLKSTFKN 260  
 DB 141 VHLRARVGEVETWTFEWNENPNDGFWEKAD-----QAVTELDV-----TARA 188  
 QY 261 AKLYGPD--VGQPRKRTAKMLKSL--KAGEVIDSVTWHHYLYNG--RTATREDPL 311  
 DB 189 IKALDFELRVGSPATAGAAVWPEFLAHVKKSGSAVDVTHYTGVDGFLDEKGVQDTKL 248  
 QY 312 NPVDLDFISSVQKVPQVVE-STRPGKWLGETSSAY 348  
 DB 249 SPSP-DAVGDVRRVRQIEASAPFGLPLFTWSTSY 285

## RESULT 12

F85875  
 Probable fimbrial usher Z3600 [imported] - *Escherichia coli* (strain O157:H7, substrain E  
 C/Species: *Escherichia coli*  
 C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C/Accession: F85875  
 R.Penna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Hiler, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potemovskis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A/Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A/Reference number: A85480; MUID:21074935; PMID:11206551  
 A/Accession: F85875  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-879 <STO>  
 A/Cross-references: GB:AE005174; NID:g12516702; PTDN:AGS7466.1; GSPDB:GN00145; UMG:Z360  
 A/Experimental source: strain O157:H7, substrain EDL933  
 C/Genetics:  
 A/Gene: Z3600

Query Match 3.7%; Score 106; DB 2; Length 879;  
 Best Local Similarity 20.5%; Pred. No. 12;  
 Matches 129; Conservative 64; Mismatches 201; Indels 234; Gaps 34;

52 VSPSFSTVTDNLATDPFPLILGSPKRTLAGISPAVYFGGKTDFLPDKKST 111  
 20 MGSQVYNAWAENBIQDSRFLEKGDPTKI-DLKRFSQGVSPG-KVNLQVQLNKOPLT 76  
 112 PERSRWQGVNODICKSIPDVBEK-----RLEWYQBLREHYQKKFNS 163  
 77 EBYDIWYASNDASKYACTLPBLVAGLQGLKEDVAKNLQWIDHDKLKPQLE----- 130  
 164 TSSRSVDVLYTFANSGDLIFGL-NALLRTADLQNNSSN-----AQLLDYC----- 211  
 131 ----GIDIK--ADLSQALVSLPQALYLETVDIMDPSPMDGDISLIDYISITQOT 182  
 212 -----SSKQYNI-SWEL-GNEPNSFL--KKADIFINGSO----- 241  
 183 RHEINGGDDSNSEISNGITGVNXXGAWRLPADWQTDYLHSGSNDDVDVINGDDTQKNWMSR 242  
 242 -----IGEDYIQLHKLKSTF-----KKAKIYGPDV 268  
 243 YWAMRALPSLKAKLGLGEDY-----LNSDIFDGFNYGGSISTDDQMLPPVLRQYAPDI 296  
 269 GQPRKRTAMLSFLKAGEVI-----DSVTMHYILNGRATATEDFLN 312  
 297 SGVATHTAKVYISQL--GRVIETQVPAGPFRIDQLGDSV-----SGTLHIREON 346  
 313 PDVLDIFISSVQKVPQVVESTRPGK--KVWL-----GETS----- 345  
 347 GQVQEDINTAMNPF-----LTRPGQVRKYLAMGRPQEMGHVEGFFSGSGEASMGCIANGM 402  
 346 SAYGGAPLLSD-----TPAA-----GFMLDL-----GLSAR 374  
 403 SLVYGA--LADEHYQSAALGVGRDLVSFGAVALFDITTSHTRLDKETAYGKGLDNGSFR 459  
 375 MGI-----EVMRQVFFGAGNYHLVDENFDPLPDVWLSLFPKLVGT--KVLMASVQGS 426  
 460 LSYSDPFDLNSRVTFAG--YRFSEENMTMSEY-LDMSDBEMRTGDKDKMTATATNQ 515  
 427 KRRKLRYVLAHCTNTDNPYKESDGLTYALNLNVTKYTL-----PYPSNKK 473  
 516 NFRDAGVSYYLVNTRHTYMDRDEQFVYVWL--SHYFVLGSIKMSISMGTGYREYDN- 571  
 474 QVDC-----YLAPLPGHGLSKSVQLNG 497  
 572 QADKGVYISLXMPMGD-----SSITSYNG 595

## RESULT 13

T38446  
 microtubule-associated protein smm4 - fission yeast (Schizosaccharomyces pombe)  
 C.Species: Schizosaccharomyces pombe  
 C.Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jul-2000  
 C.Accession: T38446; T00012  
 R.McDougal, R.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, December 1997  
 A.Reference number: Z21793  
 A.Accession: T38446  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-670 <MCD>  
 A.Cross-references: EMBL:AL009227; PIDN:CAA15832.1; GSPDB:GN00066; SPDB:SPAC2707.13c  
 A.Experimental source: strain 972h-; cosmid c2707  
 R.Yamashita, A.; Matanabe, Y.; Yamamoto, M.  
 Genes to Cells 2, 155-166, 1997  
 A.Title: Microtubule-associated coiled-coil protein smm4 is involved in the meiotic development  
 A.Reference number: Z14042; MUID:97311255; PMID:9167972  
 A.Accession: T00012  
 A.Status: translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-670 <YAM>

A.Cross-references: EMBL:AB000269; NID:g3341860; PIDN:BA31857.1; PID:g3341861  
 C.Genetics:  
 A.Gene: smm4; SPAC2707.13c  
 A.Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 9.6;  
 Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

107 KKSTPERSRWQGVNODICKSIPDVBEKRLLEWYQBLREHYQKKFNS 166  
 73 KRPSVNSR-----KKGSEINSFMEKTKALKQKRSRPSKFRSLAPLCITPIDSTPT 128  
 167 RSSVDVLYTFANSGDL-LIFGLNALRTADLQNNSSNAOLL-LDYCSSKGVNISMELGN 224  
 129 KTA--TFTSTSTNDELINFTBELSSPOTTLNLSSTLSGLDSSFMEEBEVWQVDN 186  
 225 -----BPNSEFLK-----ADIFINGSGEDYIQLHKLK--ST 257  
 187 VLQCEKKEKFPNHSKSYLKENLSLRGRDLDELKCENTALKEKIDKLEKEVEPOLT 246  
 258 FKNAKLYGPDVGPFR-KTAKMLKSLKAGEV-----DSVTMHYILNGRT 304  
 247 FLRSK--NSIEKRNFRKFLKFLAMQKEIKYLRRLQIRKIDNYKYSRSLNSKT 303  
 305 ATREDPLNPDVLD--FISVQKVPQVVESTRPGKVMGETSSAYGGAFLSDTFAA 361  
 304 PKSDNMNTQVTPSSLLGVEVSRYDL-----KQVQDITE----- 340  
 362 GFWMLDKLGSARKGIEVMRQVFFGAGN--YHLVDENFDPLPDVWLSLFPKLVGT 417  
 341 -LVKIPKPNPSEKLTISNVVRYLNIIVPSLIDQLGSLTNEFN--VAMNSTVYQELNLTK 395  
 418 VLMSVQGSRRK 430  
 396 SNNSVDGVKTRR 408

## RESULT 14

S00652  
 phosphoribosylamine-glycine lyase (EC 6.3.4.13) - fission yeast (Schizosaccharomyces pombe)  
 N.Alternate names: AIRase; aminomidazole ribotide synthetase; GARase; glycylamide rib N:contains: phosphoribosylamine-glycine lyase (EC 6.3.4.13); phosphoribosylformylglycin C.Species: Schizosaccharomyces pombe  
 C.Date: 07-Sep-1990 #sequence\_revision 28-Oct-1994 #text\_change 03-Jun-2002  
 C.Accession: S00652; T40496; T40422  
 R.McKenzie, R.; Schuchert, P.; Kibbey, B.  
 Curr. Genet. 12, 591-597, 1987  
 A>Title: Sequence of the bifunctional ade1 gene in the purine biosynthetic pathway of th A.Reference number: S00652; MUID:89003164; PMID:3502942  
 A.Accession: S00652  
 A.Molecule type: DNA  
 A.Residues: 1-788 <MCK>  
 A.Cross-references: EMBL:X06601; NID:g4903; PIDN:CAA2820.1; PID:g4904  
 R.Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.  
 submitted to the EMBL Data Library, February 1998  
 A.Reference number: Z21910  
 A.Accession: T40496  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-788 <WOO>  
 A.Cross-references: EMBL:AL021730; PIDN:CAA16823.1; GSPDB:GN00067; SPDB:SPBC4C3.02c  
 R.Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, March 1999  
 A.Reference number: Z21928  
 A.Accession: T40422  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 604-788 <SE>  
 A.Cross-references: EMBL:AL035555; PIDN:CA38600.1; GSPDB:GN00067; SPDB:SPBC405.01  
 A.Experimental source: strain 972h-; cosmid c405  
 C.Genetics:

A:Gene: A781; SPDB:SPBC405.01

C:Superfamily: Saccharomyces cerevisiae ADE5 multifunctional protein; phosphoribosylamin

C:Keywords: cyclo-ligase; purine nucleotide biosynthesis

F/5-425/Domain: phosphoribosylamine-glycine ligase homology <PGU>

F/439-767/Domain: phosphoribosylformylglycinamide cyclo-ligase homology <PFCL>

Query Match 3.7%; Score 104.5; DB 1; Length 768;

Best Local Similarity 27.7%; Pred. No. 13; Matches 70; Conservative 36; Mismatches 114; Indels 33; Gaps 11;

297 HYVNGRTATRE--DPLNDV-LDIFISSVQKQVVEST-RPGKVMLETSSAAY---- 348

424 HHALNPKRTRILLTENSISVNGNEFVORIDLVKSTRPADADIGFGFIFLQK 483

349 -GGGAPLL-SDTFAAGFMWLDKLGLSAR--MGIEVNRQVFPFAGNHYLVDENDPL--P 402

484 AGNNDPLVSVATDGVSKLIALSLNKHDTVGIDLVAMNV-----NDLVVQGAELIFL 537

403 DYMLSLPKLVGTRVLMASVQSKRRRLRYLHCTNDNPRYKSDLTLYAINLHNVTK 462

538 DYFATGSLDLKVSSTFVGVVKGCKAGCALVGETSEMPGLYHDGHYDANGTSVGAISR 597

463 YLRPYPPSNKQVQVLYLRPLGPHGLSKSVQNLGLT--KMVD---DOTLPPLEKPL 516

598 DDLPKRESFSKIDIL-----GLASDGVHNGSYLVKRIVSDLETTSSVCPMDKNV 650

517 RFGSLGUPAFSY 529

651 RLGDLSLIPRTY 663

# RESULT 15

adenylosuccinate synthetase - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 16-Jul-1999

C:Accession: F70411

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: F70411

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-432 <AOF>

A:Cross-references: GB:AE00733; NID:G2983720; PIDN:AAC07286.1; PID:G2983724; GB:AE00065

A:Experimental source: strain VFS

C:Genetics:

A:Gene: pura

C:Superfamily: adenylosuccinate synthase

Query Match 3.7%; Score 104; DB 2; Length 432;

Best Local Similarity 23.9%; Pred. No. 5.9; Matches 96; Conservative 39; Mismatches 128; Indels 138; Gaps 22;

15 LLLIGPLSPGALPPPAQADVVDLD-----FTQEPHLVSPS 55

51 ILHLPLPGILHBNHKGVAAGM-VVDLEVLHKEVKNLEKGIYKERLFISDRALHWPY 109

56 PLAVTIDANLATDPRFLILGSPK--LRTIARGLSPAYL-RFGTKTDFLIIDPKKESTP 112

110 H-----KLDSLFEKKKGIGITLKGIGAPMPKYG--RKGIKISDLKDEKRF 154

113 BERSYMQSVNODICKYGISPPDVEK-----LRLEMPQEQOLLRHGHQKKPKNSTY 165

155 ----YTLISDNLDPVK-----NICEKVFCEKFDIDINQIYEQV---RYFEFKENV- 199

166 SRSSVDVLYTPANGSGDLIFGLNALRTADL---QNNSSNAQILLDYCSSKGYNISWE 221

200 ----VDLIRFNTQKGSVLFEGAGQGLTLDVDMGTVPYTTSSNASAL-----GLSNG 246

QY 222 LGNEPNSPLKADIFING-----SOP-GEDYIQHLKLRKSTFKNAKLYG 265

Db 247 TGNPPTYP---SDAFPLGVAKATYTRVGGPPPTTEKGEGERLRL-----GGEYG 295

QY 266 PDVGQPRR---KTAKMLKSPLAGGEVIDSVTWHHYLNGRTATREDPLNP----- 313

Db 296 STGRPRRCGMIDLVLKVAQVNG-----LDGFVITKLDVLDTFDEKVCVA 343

QY 314 ----DVLDFISSVQKQVQV--VESTPRGKVMLETSSA 347

Db 344 YELDGEVIDYFPASYSSELIRKRVKVTLLKG---WKKSTGCA 381

Search completed: October 22, 2003, 20:27:38  
Job time : 46 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2003, 20:21:10 ; Search time 26 Seconds

(Without alignments)  
982.135 Million cell updates/sec

Title: US-09-759-207-2

Perfect score: 2842

Sequence: 1 MLIRSKPALPPMLALLGP.....LPASISFVIRNAVAACI 543

Scoring table: BIOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112.5	4.0	356	Y670 METUA	Q58084 methanococ
2	111	3.9	688	Y89F YEAST	P38313 saccharomyc
3	105.5	3.7	1031	TERT_EUPAE	O00939 euploides ae
4	105	3.7	670	SMW4_SCHPO	O42667 schizosacch
5	104.5	3.7	646	THIC_VIBPA	O87460 vibrio para
6	104.5	3.7	788	PUR2_SCHPO	P20772 s bifunctio
7	104	3.7	432	PURA_AQUAE	O67321 aquifex seo
8	104	3.7	716	RRP2_TAKIT	O91742 influenza a
9	103	3.6	796	PHK_CLOAB	O97163 clostridium
10	101	3.6	644	SYM_CLOAB	O97655 clostridium
11	100.5	3.5	805	GYRB_CHLPN	O92873 chlamydia p
12	99.5	3.5	837	XYNZ_CLOTM	P10478 clostridium
13	99	3.5	4594	DYHC_HUMAN	O14470 homo sapien
14	99	3.5	4644	DYHC_PAT	P18650 rattus norv
15	98.5	3.5	629	THMH_HABIN	P71366 haemophilus
16	98.5	3.5	654	NUSM_RHIST	P50367 thizopus st
17	98.5	3.5	746	RHTA_RHIME	O92395 thizobium m
18	98	3.4	716	RRP2_IATV1	P31343 influenza a
19	97.5	3.4	454	YUAE_GABEL	P52715 caenorhabd1
20	97.5	3.4	804	GYRB_CHLTR	O84193 chlamydia t
21	97.5	3.4	1314	SS22_YEAST	P23390 baccharomyc
22	96.5	3.4	595	THIC_BACHD	O9KbJ4 bacillus ha
23	96	3.4	327	XYNA_ASPAC	O59859 aspergillus
24	96	3.4	557	COX1_NEUCR	P03945 neurospora
25	96	3.4	716	RRP2_IATZ1	P13175 influenza a
26	96	3.4	1044	ITAV_MOUSE	P43406 mus muscula
27	95.5	3.4	358	VALI_BCTV	P14391 beet curly
28	95.5	3.4	620	HEWA_MEASY	P28081 measles vir
29	95.5	3.4	5255	BACH_BACLI	O68006 b bacteriaci
30	95	3.3	716	RRP2_IATKR	P13177 influenza a
31	95	3.3	772	LP1G_DROME	P11997 drosophila
32	95	3.3	2214	SORL_HUMAN	O92673 h sortilin
33	95	3.3	4644	DYHC_MOUSE	O9Jnu4 mus musculi

34	94.5	3.3	592	INV2 DAUCA	O39692 daucus caro
35	94.5	3.3	645	SYM_CLOPE	O87bJ1 clostridium
36	94	3.3	437	ERFI_XENIA	P36615 xenopus lae
37	94	3.3	804	GYRB_CHLMU	O9PKK3 chlamydia m
38	93.5	3.3	657	GRAD_TREPA	O81062 treponema p
39	93.5	3.3	766	GAPI_SCHPO	P31277 schizosacch
40	93.5	3.3	1787	UVRA_CHLMU	O9PK60 chlamydia m
41	93	3.3	449	MMI_MOUSE	O9Jkc7 mus musculu
42	93	3.3	726	CATA_ECOLI	P10229 escherichia
43	93	3.3	2733	RRPB_BUCAI	P16342 murine coro
44	92.5	3.3	455	MORF_BUCAI	P57315 buchnera ap
45	92.5	3.3	738	VAS9_SCHPO	O10145 schizosacch

## ALIGNMENTS

## RESULT 1

Y670 METUA STANDARD, PRT, 356 AA.

AC Q58084;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein M0670.  
GN M0670.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]

SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=868087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fultmann J.L., Nguyen D.,  
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.,  
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii.";  
RL Science 273:1058-1073 (1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U67514; AAB9664.1;  
DR PIR: F64383; F64383.  
DR TIGR: M0670;  
DR InterPro: IPR000051; SAM bind.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 356 AA; 41683 MW; DTB8BA2B16A92E11 CRC64;

Query Match 4.0%; Score 112.5; DB 1; Length 356;  
Best Local Similarity 21.2%; Pred. No. 0.48;  
Matches 85; Conservative 48; Mismatches 143; Indels 125; Gaps 18;

OY 126 ICKY-----GSIPVVERKLEWYQGLLRHYQKKFKSTYSRSSVD----- 171  
DB 14 IRKIKYINGKNEKDYKRLIK-----IKSHVAVETEDGTTTLAEBEEMNSKV 66  
OY 172 -----VLYTPANCGLDIFGIALMLRTADQNMSSNAQLLDVCSKGYIISWELGNEP 226  
DB 67 GALKKAIKFKAPS-----KITDL-----SNRR-VLDLCSGNGVNAIALAHYRK 109

```

Qy      217 -----NIMELGMBNSFLKKA---DIF--INGSGTG--EDYIQLKLKR----- 255
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      281 QDPAHIRSLKWDSDNKVESLRALIFNDMPFYFNKEQVDTKADGIFFRLRLRKPFKEHIN 340

```

```

Db      790 ENNAVLFI-----KJINVSRENGFKFNMMKJQTSFPLSPSKFAKYGNDSEVEQNIIVQDYC 8445

```

QY 248 QY-----HKLAKSTFKNAKLYG-----PDVCGPRKRTA-----KMLKFLKAGVYDSV 293  
 DB 846 DMIGSILPKITLALNPINILRIEGLICTLANIQTAKASMLKLLKLSL-----MANNI 899  
 QY 294 TWHYYLNRTRATREDPLNDVLDPISSVQVYFQVVESTRGKKVWLGETSSAYGGAP 353  
 DB 900 T-HYF-RKITTTEDPAKTLINKLPIISGVYKVMQCAKEKYKHFKNNLMSMIDLEVSK 955  
 QY 354 LLSDFEAGFPMWLDKGLSARNGIEVWVRQVFPAGNTHLVDENDPLDPYWLISLL----- 409  
 DB 956 IIVSVTRAFPKTL-----VCNINQDTTFGEHHY-----PDFLSTLKRFI 994  
 QY 410 -----FKLVGTATVLMASVQSKRRKR 432  
 DB 995 ELPSTKKTIFNRVCM-IILKAEKALK 1019  
 RESULT 4  
 SSM4 SCHPO STANDARD; PRT; 670 AA.  
 AC 042677  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Microtubule associated protein smm4.  
 GN SSM4 OR SPAC27D7.13C OR SPAC637.01C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NCBI Taxid=4896;  
 RX [1] SEQUENCE FROM N.A.  
 RA MEDLINE-97311255; PubMed=9167972;  
 RA Yamashita A., Matenabe Y., Yamamoto M.;  
 RT "Microtubule-associated coiled-coil protein smm4 is involved in the  
 RL meiotic development in fission yeast";  
 Genes Cells 21:155-166(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=972;  
 RX MEDLINE-21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgourou J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feldwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones W., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moutle S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbittowisch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Slimmons M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Voiclaert G., Aert R., Robben J., Grynoprez B.,  
 RA Welljens I., Vansteels B., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fytz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Manburt R., Purnelle B.,  
 RA Goffeau A., Cadieu B., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Calbert P., Aves S.J., Xiang Z., Hunt C., Moore K., Hurre S.M.,  
 RA Lucac M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovskii G.V., Ussery D., Barrett B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe";  
 RL Nature 415:871-880(2002).  
 CC -1- FUNCTION: Binds to nuclear microtubules with the effect of either  
 CC modifying their structure or function. This then promotes meiotic  
 CC nuclear division.  
 CC -1- SUBCELLULAR LOCATION: Mitotic spindle.  
 CC -1- SIMILARITY: Contains 1 CAP-Gly domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 DR EMBL, AB000269, BAA1957.1,  
 DR EMBL, AL009227, CAAL5832.1,  
 DR EMBL, AL034583, CA22580.1,  
 DR PIR, T38446, T38446.  
 DR GeneDB, Spombe, SPAC27D7.13C,  
 DR InterPro, IPR000938, CAP-Gly,  
 DR Pfam, PF01302, CAP-Gly, 1,  
 DR PROSITE, PS00845, CAP-Gly 1, 1,  
 DR PROSITE, PS0245, CAP-Gly 2, 1,  
 KW Meiosis; Microtubules; Coiled coil.  
 FT DOMAIN 23 65 CAP-Gly.  
 FT DOMAIN 209 254 COILED COIL (POTENTIAL).  
 FT DOMAIN 404 582 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 670 AA; 77105 MW; 011B7D740211B57C CRC64;  
 Query Match 3.7%; Score 105; DB 1; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 4.4;  
 Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;  
 QY 107 KKESTFEERSYQSQVNDICKYSIPDVEKRLLEWYQQLLREHYKKFNSTYS 166  
 DB 73 KRPSVVKSR-----KKSSENISNFMETKAIKOKSRREPSKERSLAPLCITPIDSSPT 128  
 QY 167 RASVDVLTTPANGSGLD-LIFGLAALLRTADLQNNSSNAOLL-LDYCSSKQYINISWEIGN 224  
 DB 129 KTA--TFYTSSTTELBELNSTEELSFDTLLNSDTSKSGLDSSFMEEEFVQVDN 186  
 QY 225 -----EPISFLKK-----ADIFINGSQGEVDYIOHLKLRK-----ST 257  
 DB 187 VLQCEKKEKFTPHSKSYLKENLKBELKRGKRDDELKCENTALKEDKLNKLEKPEPLT 246  
 QY 258 FKNAKLYGPDVGGPRR-KTAPMLKSLKAGGEV-----IDSVTHHHYLYNGRT 304  
 DB 247 FLRSK--NSIEKPRNFRREKFLKKFLAMQKIKYLRKRLQIRKIPYKYKSDSLNSKT 303  
 QY 305 ATRESPLNDVLD--ITISSVQVYFQVVESTRGKKVWLGETSSAYGGAPLSDTAA 361  
 DB 304 PKSQDQMTTQVTPSSILGVSEVSKYLDL-----KQYQVDITE----- 340  
 QY 362 GPMWLDKGLSARNGIEVWVRQVFPAGN-----YHLVDENFDPLDPYWLISLFFKLVGTK 417  
 DB 341 LVKLIPKPPFSKLTISNVNRYLNVPSLDLPQSLTNENF-----VHNSITVYQELMLK 395  
 QY 418 VLMSVQSKRRK 430  
 DB 396 SNNSSVDGVKTRR 408  
 RESULT 5  
 THIC VIBPA STANDARD; PRT; 646 AA.  
 ID THIC VIBPA  
 AC 087K60  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Thiamine biosynthesis protein thic.  
 GN THIC OR VP3027.  
 OS Vibrio parahaemolyticus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrrio.  
 NCBI Taxid=670;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=RIND 2210633 / Serotype O3:K6;

RX MEDLINE=22508454; PubMed=12620739; Yokoyama K., Uda T., Tagomori K.,  
 RA Makino K., Oshima K., Kurokawa K., Kurokawa K., Yamashita A., Kubota Y., Kimura S.,  
 RA Iijima Y., Naito T., Shingawa H., Hattori M., Iida T.,  
 RA Yasunaga T., Honda T., Shingawa H., Hattori M., Iida T.,  
 RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism  
 distinct from that of *V. cholerae*." ;  
 RL Lancet 361:743-749(2003).  
 CC -1- FUNCTION: Required for the synthesis of the hydromethylpyrimidine  
 (HMP) moiety of thiamine (4-amino-2-methyl-5-  
 CC hydromethylpyrimidine) (By similarity).  
 CC -1- PATHWAY: Thiamine biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO THE THIC FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
 CC EMBL: AF005083; BAC61290.1; -  
 DR HAMAP; MF 00089; -1.  
 KW Thiamine Biosynthesis; Complete proteome.  
 SQ SEQUENCE 646 AA; 72477 MW; FED3B87A88C237 CRC64;  
 Query Match 3.7%; Score 104.5; DB 1; Length 646;  
 Best Local Similarity 19.4%; Pred. No. 4.5; Indels 147; Gaps 23;  
 Matches 86; Conservative 56; Mismatches 153;  
 QY 43 FTQEPRLHVSPLSVITDANLAPRFLILH-GSPKLR-----TLARGLSPA 90  
 DB 61 FEENEPRVYDTS-----GYTDPDAIDLYSGPLRBEQWIERNDTEILDVSSV 112  
 QY 91 Y-----LRGG-----TKTDL--IFPKKESFEESYQSQV 122  
 DB 113 YAKRLDDELTDLRGMLPRIRAKAGCKVQLAHYARKGIVPEMEYIALREMGRAQY 172  
 QY 123 NODIC-----KYG-SIPDVEKL-----RLMEPQOULLREHYOKFE 160  
 DB 173 RQDVLTQHGSGFGANLPDITAEPRKVAAGRAIIPSNINHPSEPMIIRNLVAV 232  
 QY 161 K-----NSTYSRS--SYDVLTPRANSG--LDLIFGLNA-----LLEPTADQNMSSNA 204  
 DB 233 NANIGNSSVTSIIEBEVEKLWATRCGDTVMDSLGRNINHTREWLINSVPICITVM 292  
 QY 205 QLLLDYSSSGYINISWELNEPNSFLKADIFINSQGLGEDIYQLK--LLKSTKNAK 262  
 DB 293 YQALEKNGIAEMINWEV-----MRDTLLEQAEQGVDTYTHAGLLR----- 335  
 QY 263 LVGPDVGPARKAKMLKSFLLKAGVEYDVTW--HHYTLNGRTATREDPLNPVDLDI- 318  
 DB 336 -YVP-----MTAKRYTVIGVSGGSI--AKKCLANH-----QSFVLTTHREIC 376  
 QY 319 -FISSVQKQVQVESRPRGKRWLGETSSAYGGAPLSDTPAAGFMWLDKGLSARWGI 377  
 DB 377 EICAKYDVALSLQDGLRPGS-----VADANDBAQFAELRTIGELTKIAM 420  
 QY 378 EYVMROVFPAGN--YHVDEND 399  
 DB 421 EYDVOVIEEPGHIPIIMIKEND 444  
 RESULT 6  
 PUR2\_SCHPO STANDARD; PFT; 788 AA.  
 AC PUR2\_SCHPO; Q9UM5; -  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Bifunctional purine biosynthetic protein ADE1 [includes: (Glycinamide  
 ribonucleotide synthetase) (phosphoribosylglycinamide synthetase);

DE Phosphoribosylformylglycinamide cyclo-ligase (EC 6.3.3.1) (AIRS)  
 DE Phosphoribosyl-5-aminimidazole synthetase (AIR synthetase)].  
 GN ADE1 OR SPBC405.01 OR SPBC43.02C  
 OS Schizosaccharomyces pombe (fission yeast).  
 OS Schizosaccharomycetes; Schizosaccharomycetes;  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes  
 OC NCBI\_TaxID:4896;  
 RN [1]  
 RP -SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=89003164; PubMed=3502942;  
 RT McKenzie R., Schuchert P., Kilbey B.;  
 RT "Sequence of the bifunctional ade1 gene in the purine biosynthetic  
 RT pathway of the fission yeast *Schizosaccharomyces pombe*." ;  
 RL Curr. Gene. 12:591-597(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkwerth G., Aert R., Robben J., Grymoult B.,  
 RA Weljens I., Vansireels E., Rieger M., Schefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leinbach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode J.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;  
 RT "The genome sequence of *Schizosaccharomyces pombe*." ;  
 RT Nature 415:871-880(2002).  
 CC -1- CATALYTIC ACTIVITY: ATP + 5-phospho-D-riboseylamine + glycine = ADP  
 CC + phosphate + N(1)-(5-phospho-D-riboseyl)glycinamide.  
 CC -1- CATALYTIC ACTIVITY: ATP + 2-(formamido)-N(1)-(5-phospho-D-  
 CC ribosyl)acetamide = ADP + phosphate + 5-amino-1-(5-phospho-D-  
 CC ribosyl)imidazole.  
 CC -1- PATHWAY: De novo purine biosynthesis; fifth step.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE AIRS FAMILY.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE AIRS FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
 CC EMBL: X06601; CAA29820.1; -  
 DR EMBL; AL021730; CAA16823.1; -  
 DR EMBL; AL035655; CAB38600.1; -  
 DR PIR; S00652; S00652.  
 DR HSSP; P08178; ICL1.  
 DR GeneDB Spombe; SPBC405.01; -  
 DR InterPro; IPR000728; AIRS-related.  
 DR InterPro; IPR000115; Gars.  
 DR InterPro; IPR004733; PurM\_c1igase.



DR Pfam; PF00586; AIRS; 1.  
 DR Pfam; PF02769; AIRS; 1.  
 DR Pfam; PF01071; GARS; 1.  
 DR Pfam; PF02842; GARS; 1.  
 DR Pfam; PF02843; GARS; 1.  
 DR Pfam; PF02844; GARS; 1.  
 DR TIGRPFAM; TIGR00877; purM; 1.  
 DR TIGRPFAM; TIGR00878; purM; 1.  
 DR PROSITE; PS00184; GARS; 1.  
 DR Multifunctional enzyme; Purine biosynthesis; Ligase.  
 FT DOMAIN 1 430 GARS  
 FT DOMAIN 1 430 GARS  
 FT SEQUENCE 788 AA; 85231 MW; 0FDE64REAF9095D CRC64;  
 SQ  
 Query Match 3.7%; Score 104.5; DB 1; Length 788;  
 Best Local Similarity 27.7%; Pred. No. 6;  
 Matches 70; Conservative 36; Mismatches 114; Indels 33; Gaps 11;  
 297 HYLLNGRTATRE--DELNDV-LDIFISGVOKVQVVEST-RPGKKVWGERTSSAV--- 348  
 424 HPAALPKRTREILTYENGSGVNDNGNEFVOKIKDLVSTRPGADADIGGCGIFDLKQ 483  
 349 -GGAPPL-SDTFAAGFWMLDKLGLSAR--WGIEVVMRCVFPFAGNYHLVDENFDPL--P 402  
 484 AGMNDPLVASATDVGSKLLIALSLKHDTVGIDLVAMV-----NDLVGGAELIFL 537  
 403 DYWLILFKLVGYTKVLMASVQSKRRKRVYLHCTNTNPPRYKSGDLTVAINLHNTK 462  
 538 DYFAGSLDLKYSTSVFEGVKGKQAGALVGETSEHPRGLYHGHYDANGSTVCANVR 597  
 463 YLRLPFPFNKQVDKYLRLPGLPHGLSKSVOLNGTLT--KAVD---DQTLPLMEKRL 516  
 598 DILRKPSFSGDILL-----GLASDGVHNSNGSLVTKIVTSIDLYTSVCPMDKNV 650  
 517 RPSGLGLPAPSY 529  
 651 RLGDLSLIPTRY 663  
 Db  
 RESULT 7  
 PURA\_AOUAE STANDARD; PRT; 432 AA.  
 ID AC 067321;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DT Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase) (AdSS) (AMPSase).  
 DE PURA OR AO 1290  
 GN Aquifex aeolicus.  
 OS Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 OC NCBI\_TaxID=63363;  
 RX NCBI\_TaxID=63363;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE=98196666; PubMed=9537320;  
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.B., Overbeek R., Shead M.A., Keller M., Anjay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 aeolicus".  
 RL Nature 392:353-358 (1998).  
 CC -1- NUCLEOTIDE PLAYS AN IMPORTANT ROLE IN THE DE NOVO PATHWAY OF PURINE  
 CC NUCLEOTIDE BIOSYNTHESIS.  
 CC -1- CATALYTIC ACTIVITY: GTP + IMP + L-aspartate = GDP + phosphate +  
 CC adenylosuccinate.  
 CC -1- PATHWAY: AMP biosynthesis; first committed step.  
 CC -1- SIMILARITY: Belongs to the adenylosuccinate synthetase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 DR EMBL; AE000733; AAC07286.1; -  
 DR PIR; F70411; F70411.  
 DR HSSD; P12283; IADE.  
 DR HAMAP; MF\_00011; -; 1.  
 DR InterPro; IPR001114; Aducc synthetase.  
 DR Pfam; PF00709; Adenylosucc\_synth; 1.  
 DR Prodom; PD001188; Aducc synthetase; 1.  
 DR TIGRPFAM; TIGR00184; purA; 1.  
 DR PROSITE; PS01266; ADENYLOSUCIN SYN 1; 1.  
 DR PROSITE; PS00513; ADENYLOSUCIN SYN 2; 1.  
 DR Purine biosynthesis; Ligase; GTP-binding; Complete proteome.  
 FT NP BIND 12 18  
 FT ACT SITE 137 137 BY SIMILARITY.  
 FT ACT SITE 144 144 BY SIMILARITY.  
 FT SEQUENCE 432 AA; 48854 MW; ACA52CE913DB62E7 CRC64;  
 SQ  
 Query Match 3.7%; Score 104; DB 1; Length 432;  
 Best Local Similarity 23.9%; Pred. No. 2.8;  
 Matches 96; Conservative 39; Mismatches 128; Indels 138; Gaps 22;  
 15 LLLGLGSLSPGLPRPAQADVLDL-----FTQPLHLVPS 55  
 51 LHLPLTILHGHVAVQGM-VVDLEVLHKEVNLBEKGIYKERLFIIDRAHLVMPY 109  
 56 FLAVTTIDANLATDPRFLILGSPK--LRTLRGLSPAYL-RFGSTKDTFLIDPKKESTF 112  
 110 H-----KLDSLFEKKKGIGITLRGIPAFMYRG--RKQIRISDLKDEKRF 154  
 113 EBRSTVQVQNDICKIGSTIPDYBEK-----LRLEMPQOQLREHOKKPKSTY 165  
 155 ---YTLIEDMLDPVK-----NICEKVFCEKFDLDINQIYEBOL---RFESEKENV- 199  
 166 SRSSVDVLYTFRANGSGDLIRGNALLRTADL-----QWSSNAQLLDYCSKGVNISM 221  
 200 ---VDLRFNFTOKGSVLEBGAOCTLDVDMGYTPYVTSNANAL-----GLSNG 246  
 222 LGNEPNSFLKRDIFING-----SQL-GEYIOLHLKLRKSTPRNAKLYG 265  
 247 TQMPKRYF---SDAFGLVAKAYTTRVGECPFLTEKGESEKRLREL-----GGEYG 295  
 266 PDVGQPRF--KTAAMKSLKAGEVYDVTMHVYINGTATREDPLN----- 313  
 296 STYGFRRCGMIDLVALKYAVQVNG-----LDGVTITKLDVLDTFDEKVCVA 343  
 314 ---DVLDFISSGVOKFOV--VESTRPKKWLGETSSA 347  
 344 YELDEVIDYFPASISELIRKVPYKTLKG--WKSTXGA 381  
 Db  
 RESULT 8  
 RRP2\_IAKIT STANDARD; PRT; 716 AA.  
 ID AC 091742;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE RNA-directed RNA polymerase subunit P2 (EC 2.7.7.48) (Polymerase  
 DE acidic protein) (PA).  
 OS Influenza A virus (strain A/Kitakyushu/159/93).  
 OS Influenza A virus negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 OC NCBI\_TaxID=62478;  
 RX NCBI\_TaxID=62478;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=94406205; PubMed=9733841;  
 RA Lindstrom S.B., Hiromoto Y., Nerome R., Omoe K., Sugita S.,  
 RA Yamazaki Y., Takahashi T., Nerome K.;  
 RT "Phylogenetic analysis of the entire genome of influenza A (H3N2)  
 RT viruses from Japan: evidence for genetic reassortment of the six

CC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
CC	Bacterium
CC	NCBI TaxID=1488;
CC	SEQUENCE FROM N.A.
CC	STRAIN=ATCC 824 / DSM 792 / VGM B-1787;
CC	MEDLINE=21359325; PubMed=11466286;
CC	Moelling U., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
CC	Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,
CC	Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
CC	Bennett G.N., Koonin E.V., Smith D.R.
CC	"Genome sequence and comparative analysis of the solvent-producing
CC	bacterium Clostridium acetobutylicum";
CC	J. Bacteriol. 183:4823-4838(2001).
CC	-1- COPACITOR: Thiamine pyrophosphate (Potential).
CC	-1- SIMILARITY: BELONGS TO THE XFP FAMILY.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ).
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	EMBL; AE007645; AAK79311.1;
CC	PIR; D97065; D97065.
CC	HAMAP; MF_01403; -1.
CC	InterPro; IPR000399; Pyruvate_decarb.
CC	InterPro; IPR005593; XFP.
CC	Pfam; PF03894; XFP; 1.
CC	PROSITE; PS60002; PHOSPHOKETOLASE 1; 1.
CC	PROSITE; PS60003; PHOSPHOKETOLASE 2; 1.
CC	PROSITE; PS00187; TPP ENZYMES; FALSE NEG.
CC	KW Lyase; Flavoprotein; Thiamine pyrophosphate; Complete proteome.
CC	SEQUENCE 796 AA; 90640 MW; 488219DC778FA8 CRC64;
CC	SO
CC	Query Match 3.6%; Score 103; DB 1; Length 796;
CC	Best Local Similarity 21.6%; Pred. No 7.9;
CC	Matches 77; Conservative 51; Mismatches 114; Indels 114; Gaps 18;
CC	27 GALPRPQADVDVLDFTFGPEPLHVSPLSVTDA---NLANDPR-LILGSPKRT 82
CC	238 GWKRYFVGEDEPEPTMKLMAETLDIVBELINICKARENNDCSRPKMIVLRTPK--- 294
CC	83 LARGLSPAVLRFGSTKTDPLIFDPKKESTFEERSYQVQNDICKYGIIDPVVEKRL 142
CC	295 ---GVTGPKFV-----DGVNKGSPRAHQVLAVDRIYTEMDDLE----- 332
CC	143 EW--PYOEQLLRERHYQ--KKFKNSTYRSRSVDLYTPANCSGDLIFGLNALRTADLQ 198
CC	333 EMLKSYRPEELPDENVRLILPELBELTPKGNRMANLHAN--GGL-----LRLRLTPDR 386
CC	199 WNSSNAQLLDYCKSSKGYNISWELGNEPNSFLKPADIFINGSQIGEDYQLHKLR----- 254
CC	387 -----DYA-----VDVPTPGSTVKQMDIEIGKRYVRDVVK 415
CC	255 -KSTFKAKIYGP-----VGQPRKTKMKL--SFLKAGGEVDSVTWH-- 297
CC	416 LNEIDTRNFRIFGPDEITNSRLMAVFEQKQWLSBIEPNDEFLSNDRIYDSMISELIC 475
CC	298 -----YLLNGRTATREDPLNPDVLDIFISSVQKQVQVES--TRPGKRWIGETS 345
CC	476 EGMLEGLLNGRH-----FPASYEAFRLIIVDSMITQHGK--WLKVTYS 516

DB Methylon-1-uracil synthetase (EC 6.1.1.10) (Methionine--cRNA ligase)  
 DB (MERS)  
 GN MERS OR CAC2991.  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_Taxid=1488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE=2135925; PubMed=11466286;  
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R., Lee H.M., Dubois J., Oiu D., Hitti J., Wolf Y.I.,  
 RA Tatusov R.N., Sabathe F., Doucellet-Stamm L., Soucaille P., Daly M.J.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.,  
 RA "Genome sequence and comparative analysis of the solvent-producing  
 bacterium Clostridium acetobutylicum."  
 RT J. Bacteriol. 183:4823-4838(2001)  
 CC -1- FUNCTION: Is required not only for elongation of protein synthesis  
 but also for the initiation of all mRNA translation through  
 initiator tRNA (fMet) aminoacylation (By similarity)  
 CC -1- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +  
 L-phosphatate + L-methionyl-tRNA(Met)  
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
 CC MetG subfamily 2A.  
 CC -1- SIMILARITY: Contains 1 tRNA-binding domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@sib-sib.ch).  
 CC -----  
 CC EMBL, AB007796; AAK80932.1; --  
 DR PIR, A97268; A97268.  
 DR HAMAP, MF\_01228; fused; 1.  
 DR InterPro: IPR004495; MetG\_C-term.  
 DR InterPro: IPR002300; tRNA-synt\_1a.  
 DR InterPro: IPR001412; tRNA-synt\_1.  
 DR InterPro: IPR002304; tRNA-synt\_mec.  
 DR InterPro: IPR002547; tRNA-bind.  
 DR Pfam, PF00133; tRNA-synt\_1; 1.  
 DR Pfam, PF01588; tRNA-bind\_1.  
 DR Pfam, PF01588; tRNA-bind\_1.  
 DR PRINTS, PRO1041; TRANSYNTHEM.  
 DR TIGRfam, TIGR00398; metG; 1.  
 DR TIGRfam, TIGR00399; metG\_C-term; 1.  
 DR PROSITE, PS00178; AA tRNA\_LIGASE\_I; 1.  
 DR PROSITE, PS50886; TRSD; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;  
 KW RNA-binding; tRNA-binding; metal-binding; zinc; Complete proteome.  
 FT SITE 14 24  
 FT SITE 29 303  
 FT DOMAIN 542 644  
 FT METL 129 129  
 FT METL 132 132  
 FT METL 146 146  
 FT METL 149 149  
 FT BINDING 302 302  
 FT SEQUENCE 644 AA; 73618 MW; 30035F8B94B60A7A CRC64;  
 Query Match 3 6%; Score 101; DB 1; Length 644;  
 Best Local Similarity 18.4%; Pred.No. 8.2; indels 145; Gaps 19;  
 Matches 78; Conservative 58; Mismatches 145; indels 142; Gaps 19;  
 163 STYSSRSVDVLTYPANCGLDIFGIALNLTADLQWNSNAQLLDYSSKGVNIS--- 219  
 24 NVTYTVASALVRFRLTQGYDAFM-----LGTDEHQKQRIABDGIIRKAVV 73

QY 220 -----WELGN-EPNSFLKKADIFINGSLGEDIYOLHKLKSTFNKATLYG--- 265  
 DB 74 DEIVAGIKDLMKMMNINISYDFIRTTD-----EHVKAQVQIKVKKFDNGIITYSAY 124  
 QY 266 -----PDVGQPRRKATMLKSF--LKAGEVIDSVTHMYLYN 301  
 DB 125 EGWYCTPCESEFWETOLVDKCGPCGRVETKEAEVFKMSKADRLIKYIEHPPIQ 184  
 QY 302 GRITRE--DFLNPVDLDFISSVQKVFQVESTRGKKWLGETSAGGAPILSDT 358  
 DB 185 PESKRNEMLNFLRPGIDLCIS-----SSSPMGPIPFDE 221  
 QY 359 FAAGFMWLDKI-GLSAMGLEVWROVF--FGAGNYLVYDEN--FDLPDYMLSLF-- 410  
 DB 222 KHVIYVWDALSNVITLALGSDNDELVYKFWPADLHLVKGDIIRFTI--YVPIMLMAL 279  
 QY 411 -----KKLVGTRKVLMAVSQSKRKLKLR-----VYLHCTNTDNPYKGGDLTYAINLH 458  
 DB 280 DLPLPKQVFGHGWLL--VDGKMSKSGNVADPVVLNFEFTDVRV----- 324  
 QY 459 NVTYTLRLPYFSN-----KQYDKYLRLPLGHLGSLKSVQNLGLTKAVDDOTL 508  
 DB 325 -----YLHKEIPFGSDGFENNEIPFKTINSDLANDIG--NLVSRFA--AMIEKYFDGSIQ 375  
 QY 509 PPL 511  
 DB 376 PPV 378  
 RESULT 11  
 GYRB CHLPN STANDARD; PRT: 805 AA.  
 ID GYRB CHLPN STANDARD; PRT: 805 AA.  
 AC Q928R3; Q930Q4;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA gyrase subunit B (EC 5.99.1.3).  
 GN GYRB OR CPN0275 OR CP0484.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_Taxid=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CM1029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Baas S.,  
 RA Liner K., Weidman J., Knouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RA "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 pneumoniae AR39";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi N., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RA Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CWL029 from USA";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-  
 STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE

INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED  
DNA RINGS, INCLUDING CATEANES AND KNOTTED RINGS.  
-1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining  
of double-stranded DNA.  
-1- SUBUNIT: MADE UP OF TWO CHAINS: THE A CHAIN IS RESPONSIBLE FOR DNA  
BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE  
ENZYME FORMS AN A2B2 TETRAMER.  
-1- SIMILARITY: Belongs to the type II topoisomerase family.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AE001612; AAD18424.1;  
EMBL; AE002210; BAF38314.1;  
EMBL; AE002546; BAA98485.1;  
PIR; C66525; C66525;  
PIR; H72098; H72098;  
HSSP; P06982; IAJ6;  
PHCI-2DPAGE; Q928R3;  
TIGR; CP0484;  
InterPro; IPR003594; ATPbind\_Arpase.  
InterPro; IPR002288; DNA\_gyraseb\_C.  
InterPro; IPR000555; DNA\_gyrb.  
InterPro; IPR001241; DNA\_topoisomII.  
InterPro; IPR006171; Toprim\_dom.  
Pfam; PF00204; DNA\_gyraseb\_1.  
Pfam; PF00386; DNA\_gyraseb\_C\_1.  
Pfam; PF02518; HATPase\_C\_1.  
Pfam; PF01751; Toprim\_1.  
PRINTS; PR00418; TP12FAMILY.  
ProDom; PD149633; DNA\_gyraseb\_C\_1.  
SMART; SM00433; TOP2c; 1.  
TIGR; TIGR01059; gyrb; 1.  
TIGR; TIGR01059; gyrb; 1.  
PROSITE; PS00177; TOPOISOMERASE\_II; 1.  
Topoisomerase; Isomerase; ATP-binding; Complete proteome.  
SEQUENCE 805 AA; 90571 MW; C082DFACCEC71CCC CRC64;  
KW

Query Match 3.5%; Score 100.5; DB 1; Length 805;  
Best Local Similarity 20.4%; Pred. No. 12;  
Matches 121; Conservative 71; Mismatches 169; Indels 231; Gaps 32;

48 PHLVSPS-----FLSVITDANLATDPRFLILGSPKRLTARGSPAVL 92  
161 PLQYVSVDROGTEIVFPDPKIFSTCTD-----RSILMKRLRELAFVLRGIT 209  
93 RRGSTGTDLPFPKKESTFEERSYQSQVONODICKGSPFVVERKALEMPYQQLL 152  
210 -----IVFDDDDVDVSFDKVTFFYE-----GGQSFVS----- 236  
153 REHYOKKPKNSTSRSSVDLYTFAPCSGLHIFGLNALTADLQMNSSNAQLLDYCS 212  
237 ---YLNQKESLSS---EPYICGTRVGD-----GEIRFPAALQNSGYSLSVSYAN 284  
213 SKGYNI-SWELGEPNSFLKKADIFINGSQLGSDYIQLHKLRLKSTFFKAK-LYSPDVG 269  
285 ---NIPTRCGGTHLNGFSTALTRVIN-----TYIKAHILA-----KNNKLLATGEDL 328  
270 QPRKRTAKLAKSLKAGVYIDVTMHYTLNRTATREDFLAPDV-----LGI 318  
329 -----REGLTAVISVKNPQFEGQ--TKQDLSGSDVSVAAQVQVGEALTI 372  
319 FISS-----YQKVF-----QVVESTR-PGK-----K 338  
373 FFEENQIARMTDYK/FVAQAARAKKARELRLKRSALDSALPKLIDCLEKDEKCE 432  
339 VMLGETSSAVGGA-----PLSDTRAGFMWLDKLGSLNKGIEVMEQVFP 386

433 MYVEGSA-CGSNAQDRDRFOALIFIRKILINVEKARLOKIFONDEICTIALGCGI 491  
387 GAGVYHLYDENHFDLPYMWLSLFPKLVGVTVLM--ASVQSKRRKRV---YLACTN-- 439  
492 GADNFNL-----SKLRKRV-----IINTDADVQGSHTRLTLFFRYHNTALI 535  
440 -----TDNRYEGDLTLVAINLHNTKYLRLPYFPNKKQVDKYLRLPGRH--GLLS 490  
536 ENECVYIAQPLVY-----VSKKQDRYILSEKEMDSYILM-LGTNSSLIF 581  
491 KSV--QINGTLK-----MDDDTLPFLMEKPLRPS-----SLGLPAF 527  
582 KSTRERIRGEALSFINVILDVESFINTLEKKAIPFSEPLEMEKGIQVPLY 633

RESULT 12  
XRNZ\_CLOTM STANDARD; PRT; 837 AA.  
ID XRNZ\_CLOTM  
AC P10478  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Bnd-1,4-beta-xylanase Z precursor (EC 3.2.1.6) (Xylanase Z)  
GN (1,4-beta-D-xylan xylanohydrolase Z).  
OS Clostridium thermocellum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OK NCBI\_TaxId=1515;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIB 10682;  
RX MEDLINE=89008072; PubMed=3139632;  
RA Greifinet O., Chebrou M., Beguin P.  
RT "Nucleotide sequence and deletion analysis of the xylanase gene  
(xynZ) of Clostridium thermocellum."  
RT J. Bacteriol. 170:4582-4588 (1988).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS) OF 515-837.  
RC STRAIN=NCIB 10682;  
RX MEDLINE=95393242; PubMed=7664125;  
RA Dominguez R., Soucchon H., Spinelli S., Dauter Z., Wilson K.S.,  
RA Chauvaud S., Beguin P., Alzari P.M.  
RT "A common protein fold and similar active site in two distinct  
families of beta-glycanases."  
RT Nat. Struct. Biol. 2:569-576 (1995).  
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylan  
linkages in xylans.  
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS  
WELL AS IN OTHER C.THERMOCCELLUM CELLULOSE ENZYMES. THIS DOMAIN  
MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.  
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL  
HYDROLASES).  
CC -1- SIMILARITY: Contains 1 xynZ-type cellulose-binding (CBD) domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M22624; AAA32386.1;  
PIR; A31842; A31842.  
PDB; 1XZ; 29-JAN-96.  
PDB; 1UT2; 27-MAR-02.  
InterPro; IPR00584; CBD\_IV.  
InterPro; IPR005084; CBD\_6.  
InterPro; IPR002105; Dockerin\_1.  
InterPro; IPR002048; EF-hand.  
InterPro; IPR000801; Esterase\_put.

Interpro: IPR001000; Glyco hydro 10.  
 Interpro: IPR000379; Ser\_esterase.  
 Pfam: PF03442; CBM\_6, 1. Ser\_esterase.  
 Pfam: PF00404; Dockerin\_1, 2.  
 Pfam: PF00756; Esterase\_1.  
 Pfam: PF00331; Glyco hydro 10, 1.  
 PRINTS: PR00134; GLHYDRLASE10.  
 SMART: SM00606; CBD-IV, 1.  
 SMART: SM00633; Glyco\_10, 1.  
 PROSITE: PS00018; EF\_HAND; UNKNOWN, 2.  
 PROSITE: PS00448; CLOS\_CELLULOSE\_RPT, 2.  
 PROSITE: PS00591; GLYCOSYL\_HYDROL\_F10, 1.  
 Xylan-degradation; Hydrolase; Glycosidase; Repeat; Signal;  
 3D-structure.  
 SIGNAL 1 28  
 CHAIN 29 837  
 ACT\_SITE 645 754  
 ACT\_SITE 754 754  
 DOMAIN 328 416  
 DOMAIN 430 487  
 REPEAT 430 453  
 REPEAT 464 487  
 REPEAT 783 789  
 DISULFID 518 524  
 HELIX 525 526  
 STRAND 528 533  
 STRAND 535 539  
 STRAND 540 540  
 STRAND 543 552  
 STRAND 555 558  
 STRAND 561 562  
 HELIX 564 567  
 STRAND 570 570  
 STRAND 571 572  
 STRAND 573 573  
 STRAND 577 588  
 STRAND 589 590  
 STRAND 592 599  
 STRAND 606 609  
 HELIX 610 610  
 HELIX 615 632  
 STRAND 633 636  
 STRAND 639 643  
 STRAND 647 647  
 STRAND 649 650  
 STRAND 654 654  
 STRAND 658 663  
 STRAND 665 666  
 HELIX 667 678  
 STRAND 680 681  
 STRAND 683 688  
 HELIX 696 710  
 STRAND 711 712  
 STRAND 717 720  
 STRAND 723 725  
 HELIX 730 745  
 STRAND 746 747  
 STRAND 749 760  
 STRAND 761 762  
 STRAND 765 785  
 STRAND 787 788  
 STRAND 789 794  
 STRAND 798 798  
 STRAND 799 800  
 STRAND 804 807  
 STRAND 809 810  
 STRAND 811 812  
 STRAND 817 817  
 STRAND 819 820  
 STRAND 823 823  
 STRAND 825 834  
 HELIX 837 AA 92262 MM; DD4C9F04D12B6CD CRC64;  
 SEQUENCE

Query Match 3.5% Score 99.5; Db 1; Length 837;  
 Best Local Similarity 19.3%; Pred. No. 15;  
 Matches 63; Conservative 48; Mismatches 107; Indels 109; Gaps 16;  
 145 PROEOLINRE-----HYQKE-----KNSYRSRSVDLYTFANCGLDIFGNALDR 193  
 DB 543 PLYNSILQREFSMVYCNEMKEDALQPRONVDPFKGQDLAFARNMG-----NR 594  
 DB 194 TDLQWSSNAQLLD-----YCCKGYNISWELQNE-----PNSF 229  
 DB 595 GHTLLHNNQNPFWLITNGWNNRDSLAVKMKHITTYWTHYKGIYVMDVANECDSDGNGL 654  
 DB 230 LKADIFINGSLGSDGYIOLHKLKSTPKAKLYGP-----DYQPPRKAK 277  
 DB 655 -RSSIMWR--VIGDDYLDY-----AFRYAREDPALLFYNDYNIEDLPKSNVFN 703  
 DB 278 MKSFLKAGGEVYDSVTWHYVLTNGRTATREDFLNPDV-----LDIFISSVQKQVNES 332  
 DB 704 MKS-MKRGVYIDVGQCHFINQMSPEYIASIDONIKRYAEIGYISFTIDIRIFQS 762  
 DB 333 TRPG-----KKWLGRTSSAYGGAPLSDTPAAGFMW--LDKLGLSARMGI 377  
 DB 763 ENPATAFYQANNYKELMKICLANPN-----CTFV--MWGFTDKY----- 801  
 DB 378 EYVNRQVFRGAGNYHLVDENFDLPDY 404  
 DB 802 -TWIPGTPGYGNPLIDSNYNPKPAY 827  
 RESULT 13  
 ID DYHC\_HUMAN STANDARD; PRT; 4594 AA.  
 AC Q14204; Q92814; Q9Y4G5;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Dynein heavy chain, cytosolic (DYHC) (Cytoplasmic dynein heavy chain  
 DE 1) (DHCI) (Fragment).  
 GN DNCH1 OR DNECL OR KIAA0325.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP TISSUE=Brain;  
 RA Ohara O., Nagase T., Kikuno R., Yamakawa H., Nomura N.;  
 RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1078-1974 FROM N.A.  
 RX MEDLINE=96234671; PubMed=866668;  
 RA Valberg E.A., Grissom P.M., McIntosh J.R.;  
 RT "Mammalian cells express three distinct dynein heavy chains that are  
 RT localized to different cytoplasmic organelles."  
 RL J. Cell Biol. 133:831-842 (1996).  
 RN [3]  
 RP SEQUENCE OF 1832-1972 FROM N.A.  
 RX MEDLINE=94043467; PubMed=8227145;  
 RA Valberg E.A., Koone M.P., McIntosh J.R.;  
 RT "Cytoplasmic dynein plays a role in mammalian mitotic spindle  
 RT formation."  
 RL J. Cell Biol. 123:849-858 (1993).  
 RN [4]  
 RP SEQUENCE OF 2508-4594 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro."  
 RL DNA Res. 4:141-150 (1997).

CC -1- FUNCTION: Cytoplasmic dynein acts as a motor for the intracellular  
CC retrograde motility of vesicles and organelles along microtubules.  
CC Dynein has ATPase activity; the force-producing power stroke is  
CC thought to occur on release of ADP.  
CC -1- SUBUNIT: Consists of at least two heavy chains and a number of  
CC intermediate and light chains.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- DOMAIN: Dynein heavy chains probably consist of an N-terminal stem  
CC (which binds cargo and interacts with other dynein components),  
CC and the head or motor domain. The motor contains six tandemly-  
CC linked AAA domains in the head, which form a ring. A stalk-like  
CC structure (formed by two of the coiled coil domains) protrudes  
CC between AAA 4 and AAA 5 and terminates in a microtubule-binding  
CC site. A seventh domain may also contribute to this ring; it is not  
CC clear whether the N-terminus or the C-terminus forms this extra  
CC domain. There are four well-conserved and two non-conserved ATPase  
CC sites, one per AAA domain. Probably only one of these (within AAA  
CC 1) actually hydrolyzes ATP, the others may serve a regulatory  
CC function.  
CC -1- SIMILARITY: Belongs to the dynein heavy chain family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB002323; BAA20783.2;  
CC EMBL; U53530; AAB09727.1;  
CC EMBL; L23958; AAB16065.1;  
CC PIR; A49019; A49019;  
CC PIR; G02529; G02529;  
CC Gene; HGNC:2961; DNCH1.  
CC MIM; 600112;  
CC GO; GO:0005868; C:cytoplasmic dynein complex; NAS.  
CC GO; GO:0008577; F:dynein ATPase activity; NAS.  
CC GO; GO:000377; F:microtubule motor activity; NAS.  
CC GO; GO:0007052; P:mitotic spindle assembly; NAS.  
CC InterPro; IPR005613; AIP3.  
CC InterPro; IPR004273; Dynein\_heavy.  
CC Pfam; PF03028; Dynein\_heavy; 1.  
CC Pfam; PF03915; AIP3; 1.  
CC SMART; SM00382; AAA; 4.  
CC Motor protein; Dynein; Microtubules; ATP-binding; Repeat; Coiled coil.  
CC NON\_TER 1 1815 STEM (BY SIMILARITY).  
CC DOMAIN 1 1815 AAA 1 (BY SIMILARITY).  
CC DOMAIN 1816 2047 AAA 2 (BY SIMILARITY).  
CC DOMAIN 2128 2400 AAA 3 (BY SIMILARITY).  
CC DOMAIN 2504 2753 AAA 4 (BY SIMILARITY).  
CC DOMAIN 2847 3116 STALK (BY SIMILARITY).  
CC DOMAIN 3137 3448 AAA 5 (BY SIMILARITY).  
CC DOMAIN 3501 3730 AAA 6 (BY SIMILARITY).  
CC DOMAIN 3953 4169 COILED COIL (POTENTIAL).  
CC DOMAIN 129 150 COILED COIL (POTENTIAL).  
CC DOMAIN 403 426 COILED COIL (POTENTIAL).  
CC DOMAIN 491 514 COILED COIL (POTENTIAL).  
CC DOMAIN 1119 1200 COILED COIL (POTENTIAL).  
CC DOMAIN 1305 1321 COILED COIL (POTENTIAL).  
CC DOMAIN 3137 3223 COILED COIL (POTENTIAL).  
CC DOMAIN 3344 3448 COILED COIL (POTENTIAL).  
CC DOMAIN 3685 3748 COILED COIL (POTENTIAL).  
CC NP\_BIND 1854 1861 ATP (POTENTIAL).  
CC NP\_BIND 2172 2179 ATP (POTENTIAL).  
CC NP\_BIND 2543 2550 ATP (POTENTIAL).  
CC NP\_BIND 2885 2892 ATP (POTENTIAL).  
CC NP\_BIND 1726 1727 LH -> SD (IN REF. 2).  
CC CONFLICT 1889 1889 M -> R (IN REF. 2).  
CC CONFLICT 1973 1973 R -> N (IN REF. 2).  
CC CONFLICT 3977 3977 H -> Q (IN REF. 4).  
CC SEQUENCE 4594 AA; 527282 MW; B6F8DA7F71C169 CRC64;

Query Match 3.5%; Score 99; DB 1; Length 4594;  
Best Local Similarity 19.7%; Pred. No. 2e+02;  
Matches 113; Conservative 82; Mismatches 172; Indels 208; Gaps 27;  
QY 41 IDFFTEBFLHVSBSFLSVITDANLATDPRLLILGSKLTARGLSPATVLRGCTTD 100  
DB 1231 IDQMEQPVSVOPKRNQNDALINQKSF -PARLQVASYEPVORLLNGYMKIN 1345  
QY 101 FIPDPKESSTEEBSYVQSVNQ -DICKXGISPPDV -136  
DB 1336 MVIIEKSEA-LKOR-HKQIKRLAHVNVVSELTGQIMVNDIQKNAIVKDVLLVAVQ 1403  
QY 137 ---EE-KRLKLPYQOULLREHYOKPKNSTYSSVDVLYTFANCSGLDIFIGLN 189  
DB 1404 EWALEEFKQIRVWNTVELDV -NYQNKCR -LIRGMD 1439  
QY 150 ALKRTADIQNSSNAGLLDYCSSKGVYI -SWELENEPSFLKALDIFIN -238  
DB 1440 DLFNPKKHINSVSAMKLSPP -YKVFEDALSWB -DKLRNIMALFDVWIDVQRRW 1492  
QY 239 ---GS-----QLGEDYIOLHKLRKSTFNKAKLYGPDVGOPRRKT 275  
DB 1493 VYLSGIFPGSADIKILPEVETQROSISTEFLALMKKYSKPLVMDVNLIGVQSLRL 1552  
QY 276 AAMLKSPFKAGGEVI -DSVTWHYYLNGRTATREDPLNPDLVDF -ISSYQVVFQ 328  
DB 1553 ADLLKIKQKALGEYLERSSFPFRYFVG -DEDLLETIGNSKNVAKLQKH -1602  
QY 329 VVESTPRGKVMIGETSSAYGGGAPLSDTFAAGFMIDKGLSRMGIEVMRVVFQGA 388  
DB 1603 ---KMFAGVSSI -ILNEDNSV -VGISSEEEVWFKPP -1637  
QY 389 GNVHLVDENFDPDPYMLIFKLVGTVMASVQSKRRRLRYVLTCTNDNRYREG 448  
DB 1638 ---VSTTEPKINEMTLVEKEN -RVTLAKLAEVTEVLEFGATSIDPRTY -1686  
QY 449 DLTVALNMLHVTYKTLRLPYEPPSNQVDKY -WIDKYOQLVVLASQIAMSBNVETALSMGGGDA 1723  
DB 1687 -----IT-----1687  
QY 486 ---HGLLSK-SYQNLGLTKRVNDGTLPLPMKEPL 516  
DB 1724 APLHSVLNVEVETLVNLDVSLMEQ -PPLRRRLK 1756  
RESULT 14  
DHC RAT STANDARD; PRT: 4644 AA.  
AC P38650; Q63178; 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Dynein heavy chain, cytosolic (DHC) (Cytoplasmic dynein heavy chain)  
DE (Map 1C).  
GN DNCH1 OR DNCH1 OR DNEC1 OR MAP1C.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_Taxid=10116;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=Missar; TISSUE=Brain;  
RX MEDLINE=93376715; PubMed=7690137;  
RA Zhang Z., Tanaka Y., Nonaka S., Aizawa H., Kawasaki H., Nakata T.,  
RA Hirokawa N., "The primary structure of rat brain (cytoplasmic) dynein heavy chain,  
RT a cytoplasmic motor enzyme.",  
RT Proc. Natl. Acad. Sci. U.S.A. 90:7928-7932(1993).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Brain;  
RX MEDLINE=93264075; PubMed=7684232;  
RA Mikami A., Paschal B.M., Mazumdar M., Vallée R.B.;

"Molecular cloning of the retrograde transport motor cytoplasmic dynein (MAP 1C)."   
 CC RT Neuron 10:787-796(1993).   
 CC -1- FUNCTION: Cytoplasmic dynein acts as a motor for the intracellular retrograde mobility of vesicles and organelles along microtubules.   
 CC Dynein has ATPase activity; the force-producing power stroke is thought to occur on release of ADP.   
 CC -1- SUBUNIT: Consists of at least two heavy chains and a number of intermediate and light chains.   
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.   
 CC -1- DOMAIN: Dynein heavy chains probably consist of an N-terminal stem (which binds cargo and interacts with other dynein components), and the head or motor domain. The motor contains six tandemly-linked AAA domains in the head, which form a ring. A stalk-like structure (formed by two of the coiled coil domains) protrudes between AAA 4 and AAA 5 and terminates in a microtubule-binding site. A seventh domain may also contribute to this ring; it is not clear whether the N-terminus or the C-terminus forms this extra domain. There are four well-conserved and two non-conserved ATPase sites, one per AAA domain. Probably only one of these (within AAA 1) actually hydrolyses ATP; the others may serve a regulatory function.   
 CC -1- SIMILARITY: belongs to the dynein heavy chain family.   
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).   
 CC EMBL: D13886; BA002996.1;   
 CC EMBL: L08505; AAA41103.1;   
 CC PIR: A38905; A38905.   
 CC InterPro: IPR003593; AAA ATPase.   
 CC InterPro: IPR004273; Dynein\_heavy.   
 CC Pfam: PF03028; Dynein\_heavy; 1.   
 CC SMART: SM00382; AAA; 4.   
 CC Motor protein; Dynein; Microtubules; ATP-binding; Repeat; Coiled coil.   
 CC FT DOMAIN 1 1865 STEM (BY SIMILARITY).   
 CC FT DOMAIN 2 2097 AAA 1 (BY SIMILARITY).   
 CC FT DOMAIN 3 2178 AAA 2 (BY SIMILARITY).   
 CC FT DOMAIN 4 2554 AAA 3 (BY SIMILARITY).   
 CC FT DOMAIN 5 2897 AAA 4 (BY SIMILARITY).   
 CC FT DOMAIN 6 3187 STALK (BY SIMILARITY).   
 CC FT DOMAIN 7 3551 AAA 5 (BY SIMILARITY).   
 CC FT DOMAIN 8 3780 AAA 6 (BY SIMILARITY).   
 CC FT DOMAIN 9 4003 COILED COIL (POTENTIAL).   
 CC FT DOMAIN 10 48 69 COILED COIL (POTENTIAL).   
 CC FT DOMAIN 11 179 200 COILED COIL (POTENTIAL).   
 CC FT DOMAIN 12 453 476 COILED COIL (POTENTIAL).   
 CC FT DOMAIN 13 541 564 COILED COIL (POTENTIAL).   
 CC FT DOMAIN 14 1169 1201 COILED COIL (POTENTIAL).   
 CC FT DOMAIN 15 1229 1250 COILED COIL (POTENTIAL).   
 CC FT DOMAIN 16 1355 1371 COILED COIL (POTENTIAL).   
 CC FT DOMAIN 17 3187 3273 COILED COIL (POTENTIAL).   
 CC FT DOMAIN 18 3394 3498 COILED COIL (POTENTIAL).   
 CC FT DOMAIN 19 3735 3798 COILED COIL (POTENTIAL).   
 CC FT DOMAIN 20 1904 1911 ATP (POTENTIAL).   
 CC FT NP\_BIND 2223 ATP (POTENTIAL).   
 CC FT NP\_BIND 2553 ATP (POTENTIAL).   
 CC FT NP\_BIND 2600 ATP (POTENTIAL).   
 CC FT NP\_BIND 2935 ATP (POTENTIAL).   
 CC FT CONFLICT 1024 1025 SR -> MP (IN REF. 2).   
 CC FT CONFLICT 1772 1772 N -> D (IN REF. 2).   
 CC FT CONFLICT 2098 2098 F -> A (IN REF. 2).   
 CC FT CONFLICT 2139 2139 F -> V (IN REF. 2).   
 CC FT CONFLICT 2175 2175 D -> A (IN REF. 2).   
 CC FT CONFLICT 2185 2185 K -> Q (IN REF. 2).   
 CC FT CONFLICT 2366 2366 L -> V (IN REF. 2).   
 CC FT CONFLICT 2382 2382 T -> S (IN REF. 2).   
 CC FT CONFLICT 2463 2463 G -> A (IN REF. 2).   
 CC FT CONFLICT 3219 3219 A -> D (IN REF. 2).   
 CC FT CONFLICT 4131 4131 R -> K (IN REF. 2).

FT CONFLICT 4366 4366 F -> S (IN REF. 2).   
 FT CONFLICT 4511 4511 A -> G (IN REF. 2).   
 SQ SEQUENCE 4644 AA; 532240 MW; 8C6ABDBED875D82 CRC64;   
 Query Match 3.5%; Score 99; DB 1; Length 4644;   
 Best Local Similarity 19.5%; Pred. No. 2e+02;   
 Matches 109; Conservative 85; Mismatches 188; Indels 178; Gaps 25;   
 QY 41 LDFTQEPHLVSPFLSVTTIDANLATDPRLILSGPKRTLARGSLPAYLRFQRTKD 100   
 DB 1341 IDQKEPWPVSVQPKRLRMDIDGLNQLKIF-----PARLRQVSYEFVQRLGVMKIN 1395   
 QY 101 FLIDPKKESFEERSVQGVNQ-----DICKXGSPDPV----- 136   
 DB 1396 MLVIELKSEA-LKQR-HWKQLMKRLHVMNVVSELTQIWDVDLQKREAIYKDVLLVAQG 1453   
 QY 137 -----ER---KLRLWPQEOQLLREHYQKFKNSTYSSVDVLYTPANGSLDIFGLN 189   
 DB 1454 EMALPEFLKQIRVWNTYELDV--NYQNKCR-----LIRGWD 1489   
 QY 190 ALFTADLQWSSNAQLLDYCSKGYNI-----SWELEPNPSFLKADIFIN----- 238   
 DB 1490 DLFNKVKRHNSVSAKLSFY-----YKVEEDALSWE--DKLNRIVALFDVWIDVQRW 1542   
 QY 239 -----GS-----QGEVDYQHLKLRKSTFNKATLQGVQGVQPRKRT 275   
 DB 1543 VYLEGIFGTSADIKHLPLVETORFQSTSTFELAMKRVKSPVMDVLTNOGVRSLERL 1602   
 QY 276 AKMLKSLKAGGEVIT--DSVTWHYLYNGRTATREDPLNDVDLIF-----ISSQVKVFQ 328   
 DB 1603 ADLLGKQKQKGLGELVLEHRSRPFYFVG-----DEDLLEITIGNSKVAQLQHF- 1652   
 QY 329 VVESTRPQKRWLGITSSAVGGAPLLSDTPAAGFWMLDKLGASARNGIEVWQVFEFGA 388   
 DB 1653 -----KKMFAGVSSIIINEDSSVY-----LGSISSREGEEVWFPTP----- 1687   
 QY 389 GNHYLVDENDPLPDVWLSLFFKLVGTQKYLMAVQGSKKRKLRYVLAHCTNDPRY--- 445   
 DB 1688 -----VITHPKINEMTLVLEKEM--RVTLAQLAESVTEVAIFPKATSIDNTYITW 1739   
 QY 446 ---REGQTLVAINL---HNVTKYLRLPFPNSKQVDKYLRLPFGGLSLKSYQLNGLT 499   
 DB 1740 IDKQADLVLSAGIANSSEVENALNSVGGGN-----VGPLQSVLSNVE---VT 1786   
 QY 500 LKAVDDQTL---PPLMEKPL 516   
 DB 1787 LNVLADSVLMEQPLRRKL 1806   
 RESULT 15   
 T3MH\_HAEIN STANDARD; PRT; 629 AA.   
 ID T3MH\_HAEIN STANDARD; PRT; 629 AA.   
 AC P71366;   
 DT 16-OCT-2001 (Rel. 40, Created)   
 DT 16-OCT-2001 (Rel. 40, Last sequence update)   
 DT 16-OCT-2001 (Rel. 40, Last annotation update)   
 DE Putative type III restriction-modification system HindVIP enzyme mod   
 DE (EC 2.1.1.72) (HindVIP methyltransferase) (M.HindVIP).   
 GN H11056.   
 OS Haemophilus influenzae.   
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;   
 OC Pasteurellaceae; Haemophilus.   
 OX NCBI\_TaxID=727;   
 RN [1]   
 RP SEQUENCE FROM N.A.   
 RC STRAIN=Rd / KW20 / ATCC 51907;   
 RX MEDLINE=95350630; Pubmed=7542800;   
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,   
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,   
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,   
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,   
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,   
 RA Uettermack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

RA Fine, L.D., Fritschman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm, C.L., McDonald, L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.,  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Science 269:496-512(1995).  
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-  
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.  
CC -1- SUBUNIT: CONTAINS TWO DIFFERENT SUBUNITS: RES AND MOD. MOD IS  
CC N HOMOTETRAMER (BY SIMILARITY).  
CC -1- SIMILARITY: WITH OTHER TYPE III MOD PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC  
CC EMBL: U32786; AAC22721.1; -  
CC PIR: C64180; C64180  
DR REBASE: 3701; M.HINDORF1056P.  
DR TIGR: H11056;  
DR InterPro: IPR001091; CNA\_Mettransf.  
DR InterPro: IPR002295; D2IN6\_Mtase.  
DR InterPro: IPR002941; N6/N4\_Mtase.  
DR InterPro: IPR002052; N6\_Mtase.  
DR Pfam: PF01555; N6\_N4\_Mtase; 1.  
DR PRINTS: PR00506; D2IN6MTFRASE.  
DR PRINTS: PR00508; S2IN4MTFRASE.  
DR PROSITE: PS00092; N6\_MTASE; 1.  
KW Hypothetical protein; Transferrase; Methyltransferase; DNA-binding;  
KW Restriction system; Complete proteome  
SQ SEQUENCE 629 AA; 71845 MW; 93ADAD909DB41B84 CRC64;

Query Match: 3.5%; Score 98.5; DB 1; Length 629;  
Best Local Similarity 21.8%; Pred. No. 12;  
Matches 67; Conservative 56; Mismatches 107; Indels 77; Gaps 19;

QY 99 TDFL-IPDPKESTFEERYNOSQVODICKYGISPPDVEEKLRLWPYQ-----EQL 150  
DB 242 TEFIVFSKNRRRLFLNPLVYQKTEVNEFIKNY-----EDSGKSWKYQVLIIDLGSKI 293  
QY 151 LIREHYOKKFKKSTYSRSVDVLYTFANCSGL--DLIFG--LNALLRTADLQ----- 198  
DB 294 LHEE--KQGFKYHFNAMQTSIVKFSQDQNLKSKILLIYESHKYYRTTMAOSSIRSKII 351  
QY 199 ---WNSSNAQLLDYCSSKQYN-----ISMELGNEPNSFLKADIFING----- 239  
DB 352 EDLYSIKNGIVIEIYIPQKKNAGMLIEVFYNASNK--DMFMFLSDMLIKKKKYFYLOKV 410  
QY 240 SOLGSDYIQHLGLKKS---TFKNAKLYQPDVGOPRRTAKMLKSLKAGSGVIDSVTWH 296  
DB 411 NTLMDQ-IQYNNLNKEGGYIDPKNGK-----KPEALLRIIDMTTKGDIYLD----- 457  
QY 297 HYLL-NGRTATREDFNPDVLDI---FTS--SVQKFOVESTSPG--KKY-MLGSTSS 346  
DB 458 -YHLSSGTTAAVAHAKMROYIGIBQMDYLETTLAVELKVKVIDGEOGGISKAVNMGGGBF 516  
QY 347 AVGGAP 353  
DB 517 VYABLAP 523

Search completed: October 22, 2003; 20:24:56  
Job time : 29 secs